



De Sousa, Gabriela (2021) *Neural correlates of multisensory decision making in humans*. PhD thesis.

<http://theses.gla.ac.uk/82474/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses
<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk



University of Glasgow | Institute of Neuroscience
& Psychology

Neural correlates of multisensory decision making in humans

Gabriela De Sousa

BSc Psychology, MSc Cognitive Neuroscience

Submitted in fulfilment of the requirements for the Degree of Doctor
of Philosophy

Institute of Neuroscience and Psychology
College of Medical, Veterinary and Life Sciences
University of Glasgow

April 2021

Abstract

In everyday life we are constantly required to make decisions about things that we perceive in order to interact with our environment and perform everyday tasks. Perceptual decision making is the process by which sensory evidence is collected and accumulated towards one of two or more possible choices. This is an inherently a noisy process, and decisions made need to optimally trade-off between speed and accuracy, as well as combine complementary evidence of more than one sensory type (here, sound and vision). While great progress has been made in understanding the neural correlates of unisensory perceptual decision making, relatively little is known about the enhancements or changes to this process that result from the integration of more than one modality of information.

The current thesis presents empirical findings from three studies that sought to provide a more complete characterization of multisensory decision making using electrophysiological and diffusion modelling methods. Specifically, **Study 1** (Chapter 2) investigates the temporal evolution of audiovisual decision making and compares whether early sensory integration or late post-sensory decision processing of visual evidence is enhanced in the presence of complementary auditory information. We recorded EEG measurements from human subjects during performance of a face versus car categorisation task. On some trials, participants were presented with images alone, while in others we simultaneously presented sounds of the same object category (i.e. speech and car sounds). Responses were more accurate and slower during audiovisual trials, and both accuracy and response time scaled with sensory evidence. Neural activity discriminating between face and car trials was observed peaking shortly before the time of response in a fashion that mirrored the process of evidence accumulation. This interpretation was confirmed using a neurally-informed drift diffusion model. Further, we found that trial-by-trial changes in behaviour could be predicted by neural activity within this model. Topographical representations of these signals revealed a prominent centroparietal cluster of activity.

Leading on from this, **Study 2** (Chapter 3) modified a continuous version of the dot motion discrimination task to include sound motion and audiovisual motion trials. Participants received no obvious sign as to the start of a coherent motion period, which therefore prevented visually-evoked potentials and provided an

unimpeded observation of evidence accumulation activity from the beginning of the trial to the point of decision. In doing so, we sought to further understand the enhancement of evidence accumulation activity during audiovisual trials. We focused on the same centroparietal cluster that we had observed in the previous chapter, and that was highlighted in the original study by Kelly and O'Connell (2013). Participants missed significantly fewer trials with audiovisual motion. Activity clearly increased at a steady rate from around 200/300ms post-stimulus onset, up until the point of response, in a pattern again mirroring evidence accumulation. We found that this activity was again enhanced during audiovisual trials compared to visual-only trials, with greater rates of increase in activity. Activity also peaked at a slightly higher level shortly before the time of response. These findings supported those of the Chapter 2 in that the presence of complementary auditory information enhanced the decision making process.

Finally, we asked whether oscillatory patterns within the EEG signals may offer additional insights into the neural representations of multisensory decision making. We extended the investigation of neural signals collected in Chapter 3 using the continuous dot motion discrimination task by decomposing the original broadband signal into its component frequencies, here focusing on beta, gamma, and high-gamma activity. We compared the rate of change in power between sensory conditions leading up to the time of response, as well as shortly after. While we did find interesting modulations in power relating to specific sensory conditions within the task, including a pattern of desynchronization that may suggest input from premotor structures in the embodiment of the decision, we did not find the same robust modulation in evidence accumulation by sensory condition that we had observed in the previous chapters. However, we could clearly see gradual changes in power that seems to reflect evidence accumulation.

Together, our results reveal novel insights into the neural representations of multisensory decision making in the human brain and point to new research directions that may uncover more about the neural underpinnings of audiovisual decision making. It also suggests further study of related activity such as decision confidence, or the embodiment of evidence accumulation within premotor areas.

Table of Contents

Abstract	2
Acknowledgement	6
List of Tables	7
List of Figures	7
List of publications	8
Author's Declaration	9
Abbreviations	10
1 General Introduction	11
1.1 Unimodal perceptual decision making.....	12
1.2 Multisensory integration	13
1.3 Multisensory decision making, optimisation, and facilitation.....	14
1.4 Aims for the thesis	17
2 Chapter 2. Audiovisual sensory evidence enhances post-sensory decision processing	19
2.1 Summary	19
2.2 Introduction	20
2.3 Methods	23
2.3.1 Participants	23
2.3.2 Stimuli and task.....	24
2.3.3 Behavioural analysis.....	27
2.3.4 EEG data acquisition and pre-processing.....	28
2.3.5 EEG data analysis	29
2.4 Results	36
2.4.1 Behavioural results	36
2.4.2 Temporal impact of auditory evidence on visual representations.....	38
2.4.3 Neurally-informed cognitive modelling.....	43
2.4.4 Neurally-informed model outperforming behaviourally constrained model.....	46
2.5 Discussion	48
2.5.1 Early and late accounts of multisensory decision making	48
2.5.2 Using neurally-inspired models to understand decision making	50
3 Chapter 3. Centroparietal positivity reflects audiovisual evidence accumulation.....	54
3.1 Summary	54
3.2 Introduction	55

3.3	Materials & Methods	60
3.3.1	Participants	60
3.3.2	Stimuli and task.....	60
3.3.3	Behavioural analysis.....	66
3.3.4	EEG data acquisition	67
3.3.5	EEG data pre-processing.....	68
3.3.6	EEG data analysis.....	68
3.4	Results	71
3.4.1	Behavioural results	71
3.4.2	EEG results	75
3.5	Discussion.....	87
4	Temporal characterisation of oscillatory activity during audiovisual perceptual decision making.....	95
4.1	Summary	95
4.2	Introduction	96
4.3	Materials & Methods	101
4.3.1	Participants	101
4.3.2	Stimuli and task.....	101
4.3.3	EEG data acquisition	102
4.3.4	EEG data pre-processing.....	103
4.3.5	EEG spectral analysis.....	103
4.3.6	Statistical analysis.....	105
4.4	Results	108
4.4.1	Behaviour	108
4.4.2	Spectral analysis.....	108
4.5	Discussion.....	118
5	General Discussion	126
5.1	Overview.....	126
5.2	Key findings	127
5.3	Limitations and future directions	130
5.4	Conclusion	132
	List of References	133

Acknowledgement

First and foremost, I would like to thank my supervisor, Professor Marios Philiastides, for his advice, guidance, and support throughout my PhD. I am very grateful to have been a part of your lab and to have had such a range of opportunities to grow my skills and experience as a researcher. I especially want to say thank you for your continued support as I worked to complete my thesis; your feedback gave me the confidence I needed to keep going.

I would further like thank my supervisor, Professor Christoph Kayser, for your guidance and analytical advice. Thank you also to Dr Esther Papies, Professor Gregor Thut, and Dr Guillaume Rousselet, who gave me valuable feedback as I progressed through my studies and helped to keep my research on the right path.

I was extremely lucky to be part of a lab with such bright and encouraging people: Andrea, Elsa, Ema, and Filippo, your experience was invaluable as I was finding my feet as an early PhD student. Jessie and Sabina, thank you for passing on so much of your wisdom. I especially want to thank Léon for the countless hours spent on our collaboration, and your unending patience.

Thank you to the wonderful colleagues that I met in the department: Gemma, Maisy, Gaby, and Greta, our many conversations made the tough days so much easier. Thank you to all the other PhD students I had the pleasure of meeting and sharing this experience with.

To my family, especially my mum, dad, sister, and brother; your constant faith in me means so much, even though I couldn't be home as much as we'd hoped. To my Glasgow family; Gemma, Kirstin, Katy, Caitlin, Heather, and Lauren. I am so lucky to have found friends in such kind, strong, and intelligent women. To Euan especially, thank you for your unwavering support and patience. I certainly would not have made it to the end without you all.

I would not like to thank SARS-CoV-2 or any of its variants for anything, however I would like to thank all of the incredible members of the research community and the NHS for your tireless efforts during a truly *unprecedented* time.

List of Tables

Table 4.1 Specific timings of slope and peak power per period of interest and frequency band.	109
-----------------------------------------------------------------------------------------------------------	-----

List of Figures

Figure 2.1 Experimental paradigm.	23
Figure 2.2 Behavioural performance.....	37
Figure 2.3 Stimulus-locked face-vs-car discrimination analysis.	39
Figure 2.4 Response-locked face-vs-car discrimination analysis.....	42
Figure 2.5 Neurally-informed cognitive modelling.....	44
Figure 3.1 Behavioural task design.....	63
Figure 3.2 Behavioural results of audiovisual motion task.	72
Figure 3.3 Topography of ramping activity, collapsed across conditions.....	75
Figure 3.4 Temporal profile of stimulus-locked ERP from CPP sensors.	76
Figure 3.5 ERPs locked to the time of response, per sensory condition.	80
Figure 3.6 Robust linear regression predicting miss rate with ERP slope steepness (i.e. evidence accumulation rate).....	84
Figure 4.1 Temporal evolution of power estimates, relative to the time of response.	109
Figure 4.2 Spectral and statistical analysis results of period of interest leading up to the time of response ('pre-response' period).....	112
Figure 4.3 Spectral and statistical analysis results of period of interest at and shortly after the time of response ('post-response' period).	115

List of publications

Franzen, L., Delis, I., De Sousa, G., Kayser, C. and Philiastides, M.G., 2020. Auditory information enhances post-sensory visual evidence during rapid multisensory decision-making. *Nature Communications*, 11(1), pp.1-14.

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Abbreviations

A	Auditory (trials)
AV	Audiovisual (trials)
CPP	Centroparietal positivity
DLPFC	Dorsolateral prefrontal cortex
EA	Evidence accumulation
EEG	Electroencephalography
ERP	Event-related potential
fMRI	Functional magnetic resonance imaging
LIP	Lateral interparietal (area)
LRP	Lateralised readiness potential
mPFC	Medial prefrontal cortex
PDM	Perceptual decision making
V	Visual (trials)
VEP	Visually-evoked potential

1 General Introduction

Perceptual decision making is the process whereby humans and other animals identify and make behavioural decisions about their surroundings in a way that changes with their environment. Sensory information is collected and evaluated until there is sufficient sensory evidence in favour of an option, in order to decide how best to respond and prepare relevant motor actions (Hauser and Salinas, 2014). Neuroscientists investigate this process in an attempt to understand how we perceive and react to our environment using an interconnected hierarchy of cortical areas. In order to explore its topographical and temporal properties, a behavioural task will challenge participants to discriminate between finely tuned stimuli that may range in difficulty, causing differences in reaction time (RT) within each trial. Greater difficulty levels also allow longer decision times, creating a better opportunity to measure cortical differences through methods such as electroencephalography (EEG).

Recent literature has begun investigating the effects of the bimodal sensory presentation of stimuli on the resulting decision process, with examples of both informative and uninformative stimuli. It is likely that the presence of additional, task-relevant information could facilitate the decision process, as related information gathered from multiple sensory modalities is likely to be integrated to better perform a task (Shi and Müller, 2013). However, the neural underpinnings of these changes require further exploration in order to understand what cortical changes are causing this effect.

The following introduction attempts to summarise recent literature exploring the role of multisensory integration within perceptual decision making. First, it will establish the current understanding of perceptual decision making. Further, it will explore evidence of multisensory integration within early sensory cortices. Finally, it will review recent examples of multisensory perceptual decision making.

1.1 Unimodal perceptual decision making

A simple unisensory task, for example visual, is normally used in order to observe perceptual decision making in participants. Differences in accuracy and RT can be used to measure performance differences between groups, or included in Bayesian models such as the drift diffusion model (DDM; Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006) which take a number of parameters into account, in an attempt to predict responses and better understand the decision process. These strategies are often implemented in combination with a method of neuroimaging such as EEG or functional magnetic resonance imaging (fMRI). These have been successfully used to reveal a hierarchical procedure of activation during perceptual decision making (Gold and Shadlen, 2007; Ratcliff, Philiastides and Sajda, 2009; Ding and Gold, 2010).

Traditionally, perceptual decision making tasks have presented visual stimuli quickly and with immediate onset and offset times, causing them to only flash briefly onscreen. This allows more completed trials and some control of difficulty, and tasks of this manner have revealed many insights into the decision making process. For example, random dot motion (RDM) discrimination tasks have suggested the role of the lateral inferior parietal cortex in systematic evidence accumulation (Shadlen and Newsome, 1996; Ho, Brown and Serences, 2009).

However, sudden stimulus presentations increase noise from basic sensory processing and could be facilitating detection performance in some cases. It may be better to try to mask the onset of a stimulus. O'Connell, Dockree, and Kelly (2012) used a flickering annulus that gradually changed in contrast, and asked participants to detect when they noticed the stimulus begin to dim, while recording using EEG. This task caused greater variation in RTs, however almost all incidences were noticed by participants. Analysis revealed a centroparietal positivity (CPP) event-related potential (ERP) that scaled in strength as evidence was accumulated, regardless of whether detection required a button press or mentally keeping count of the number of changes. CPP changes matched those of motor left hemisphere beta preparatory activity for button pressing, suggesting the role of CPP as a supramodal decision variable component. A further study used gradual presentation of RDM detection trials, again finding

that CPP increased in activity in proportion with the rate of evidence accumulation, suggesting the role of CPP as a supramodal decision variable (Kelly and O'Connell, 2013). These two experiments also avoided problems of peak responsivity with sudden stimulus onset by using this gradual onset, continuous monitoring design. This CPP component has been linked to the P300 component found in previous literature as being highly involved in evidence accumulation and the decision process (Rohrbaugh, Donchin and Eriksen, 1974; Duncan-Johnson and Donchin, 1982; Kelly and O'Connell, 2015). This literature suggests the potential of gradual onset, continuous monitoring designs in order to better explore the neural activity of perceptual decision making.

1.2 Multisensory integration

It is the multisensory integration of information between sensory cortices that make up the early stages of multisensory decision making and the facilitatory effects it may have. Romei, Murray, Cappe, and Thut (2009) found greater incidences of phosphenes caused by transcranial magnetic stimulation when accompanied by naturally threatening 'looming' sounds, suggesting that they had increased excitability of the visual cortex. In an audiovisual fMRI experiment, participants tended to respond superadditively and faster to bimodal stimuli compared to unimodal stimuli (Brang et al., 2013). They also found direct pathways between the primary visual and auditory cortices, with a relationship between anatomical connectivity and multisensory processing. Kayser, Petkov, Augath, and Logothetis (2007) used fMRI to reveal the modulation of early auditory cortex by presentation of visual scenes. Lange, Christian, and Schnitzler (2013) observed audiovisual oscillation synchronisation in a speech task using magnetoencephalography (MEG), where Broca's area and the auditory cortex exhibited coupling during congruent stimuli. Mercier et al. (2013) also observed oscillatory activity resetting of the visual cortex, as modulated by input from auditory stimuli. They also found evidence of auditory ERPs within visual areas themselves.

The above clearly demonstrates that a large pattern of interconnectivity exists within early sensory cortices that serves to modulate the activity of one area based on the simultaneous information gained from others. This likely has a key

role when performing perceptual decision making tasks with multimodal information.

1.3 Multisensory decision making, optimisation, and facilitation

A number of recent studies have begun exploring the use of bimodal stimulus presentation during perceptual decision making tasks. The primary findings have been of behavioural facilitation resulting from bimodal sensory evidence. Raposo, Sheppard, Schrater, and Churchland (2012) found that both humans and rats could perform event stream rate estimations at near statistically optimal levels when both visual and auditory stimuli were presented, compared to unimodal presentation. Chen, Huang, Yeh, and Spence (2011) asked human participants to complete a Gabor patch discrimination task, with some trials also including a simultaneous audio cue. Both discrimination and detection performance were enhanced in bimodal trials compared to unimodal. Interestingly, the audio component was uninformative of the discrimination task. These behavioural advantages seem to cross species, with Kulahci, Dornhaus, and Papaj (2008) finding that bees trained to discriminate using two sensory modalities could learn to identify rewarding flowers faster than those trained unimodally.

Literature exploring cortical processing underlying these behavioural changes is an emerging area of interest. In an EEG study by Stekelenburg and Vroomen (2012), participants completing a biological motion perception discrimination task showed early N1 and P2 auditory ERP component suppression during trials with audiovisual stimuli. Naci, Taylor, Cusack, and Tyler (2012) presented unimodal or bimodal audiovisual animal stimuli, during EEG. Early superadditive activity was found in the anterior temporal cortex and inferior prefrontal cortex during bimodal tasks. Activity in the posterior occipital cortex appeared later, suggesting top-down feedback processes during multisensory decision making. Otto & Mamassian (2012) used an audiovisual decision task where either modality was sufficient in order to make a decision, and found that evidence accumulation occurred separately for each sensory modality, before integration with a logical operator.

Some studies note changes in oscillatory activity during bimodal perceptual decision tasks. Gleiss and Kayser (2013) asked participants to perform a two-interval RDM discrimination task while a white noise sound played. The sound was either stationary through the whole trial or moved coherently, or incoherently, horizontally with the visual stimulus. Audiovisual modulation of low frequency and alpha band activity, particularly at 300ms post-stimulus, was predictive of a perceptual multisensory benefit, in that the additional presence of a sound facilitated visual processing required to detect motion. A behavioural improvement was most apparent during moving sound trials. A following study by Gleiss and Kayser (2014), this time using Gabor patch visual detection, played either continuous noise or transient sounds and recorded electrocortical activity using EEG. Both transient and continuous sounds resulted in increased task performance, which was linked to reduced alpha-band power as audio noise increased.

These results, and those of the former experiment, suggest that alpha-band oscillatory changes could be monitored as an indirect measure of visual cortex sensitivity during bimodal decision tasks, or that attention plays a key role in processing of these stimuli. There are also the suggestions of a number of additional effects on perceptual decision making processing involving simultaneous bimodal presentation, such as increased RTs and early feedback from higher processing areas to sensory cortex.

Statistical models of behavioural data have become a more common method of explaining the underlying processes and features of perceptual decision making such as optimisation. As described previously in research by Raposo et al. (2012), several other studies have discussed measuring reaction time and error rate during a two-alternative forced-choice task, in order to model the speed-accuracy trade-off (SAT) of decisions. SAT describes how response behaviour may be strategised as a balance between fast, less accurate decisions and slower, more accurate decisions, in order to maximise reward. The most popular statistical models, such as the DDM, attempt to describe the decision process as the gradual accumulation of evidence until a decision boundary or threshold is reached, and a final decision made. An amount of noise is also accounted for, and the placement of the decision bounds may change depending on the SAT of a

particular decision. If a faster decision is required to maximise reward, the bounds may be closer to the midpoint, whereas if accuracy is a priority they may be further away, meaning the decision takes longer but more evidence has been accumulated before this point (Bogacz *et al.*, 2006; Ratcliff and McKoon, 2008; Krajbich and Rangel, 2011). SAT has also been linked to increased baseline activity in the pre-supplementary motor area and the dorsolateral prefrontal cortex during speed-emphasised tasks (Bogacz *et al.*, 2010), which are areas related to preparation and association more so than sensory processing.

Some statistical models that can predict the performance enhancements resulting from multisensory information and integration were only able to do so when both stimulus presentation time and response time were kept constant throughout the task. While the findings were interesting, other studies such as that by Drugowitsch, DeAngelis, Klier, Angelaki, and Pouget (2014) criticised this model, stating that variable reaction times should also be taken into account to maintain ecological validity. In real-world settings, participants make decisions with varying response speed, often waiting until they have sufficient information within the context before making a choice. While a model by Clark and Yuille (1990) describe participants as optimising their decisions to combine information across modalities when presentation and reaction times were fixed, Drugowitsch *et al.* found that data with varying reaction times was found to be suboptimal by the same criteria. However when using their own model that took varying reaction times into account, responses were instead found to be near statistically optimal, meaning that information gained in multisensory trials was equal to the sum of that gained from each unisensory stimulus. This suggested that participants were able to combine information across sensory modalities in a way that gave the highest likelihood of reward, and that SAT played a role in this. However, recent studies have found evidence suggesting that this total summation of sensory information is imperfect. Carland *et al.* (2016) observed the effects of brief motion pulses on a variant of a constant-coherence motion discrimination task when pulsed trials were presented in blocks or were interleaved. Pulses slowed responses during later interleaved trials when participants slowed their decision policy, which the authors suggests reflects a growing urgency signal for decisions as the task progresses, that is more in line with an urgency-gating model. In this model, it is possible for information to

‘leak’, which the DDM does not account for, suggesting that the DDM may not model perceptual decision making behaviour completely.

1.4 Aims for the thesis

As this chapter has highlighted, considerable prior research exists that has characterised the neural underpinnings of perceptual decision making, but more recently the need to investigate similar mechanisms within multisensory decision making has emerged and begun to take shape.

Our first study (Chapter 2) aims to explore the temporal evolution of decision making activity by building on an established behavioural and analytical paradigm, and observing how neural representations of visual evidence are enhanced when complementary auditory evidence is provided in tandem. In short, we collected EEG measurements from human participants as they made speeded categorisations of images as of faces or cars, however we additionally played speech or car sounds in half of the trials. Using a single-trial linear discriminant analysis, we found that neural signals of both trial types mirrored a pattern of decision-related evidence accumulation, gradually rising to a peak shortly before the time of response. We showed that the rate of evidence accumulation increased significantly during audiovisual trials, and this was corroborated by a neurally-informed drift diffusion model which also found that behaviour could be predicted using this activity.

Our second study sought to investigate the change in evidence accumulation further, and used a version of the classic random dot motion discrimination task (Ratcliff, 1978) that had been modified to present a seamless transition between incoherent and coherent motion periods (Kelly and O’Connell, 2013). We then modified this further by adding an auditory element, allowing us to compare evidence accumulation rates and peak activity between visual, auditory, and audiovisual trials. We again showed an increase in the rate of evidence accumulation during audiovisual trials compared to those with visual evidence only, indicating an enhancement of this process that coincides with the complementary information provided. We then asked whether further

information may be contained within oscillatory patterns of activity within the broadband signal that we had already analysed. Therefore, for our third results chapter we conducted a reanalysis of the data from our second study, decomposing the signal into beta, gamma, and high-gamma fluctuations in power. The primary goal of this approach was to explore whether more could be learned from the neural signals of enhanced evidence accumulation that we had already seen in our previous chapters. While they generally followed this pattern, we did not observe the same clear differences between sensory conditions that we had seen before. However, the results of this chapter and those before indicated there may be potential in studying the relationship between multisensory decision making and increased confidence, or whether there is more to be learned about the embodiment of evidence accumulation within premotor areas by using a similar paradigm.

2 Chapter 2. Audiovisual sensory evidence enhances post-sensory decision processing

Note on contributions to work contained within the following chapter: I played a central role in experiment planning and implementation, data collection, and data analysis that gave rise to the results included in sections 2.4.1 and 2.4.2, however 2.4.3 on was developed further and completed by colleagues as part of work towards our publication in *Nature Communications* (Franzen *et al.*, 2020), in which I am joint-first (i.e. equally contributing) author. This chapter as presented here is an earlier version of that work.

2.1 Summary

Multisensory decision making requires the combination of different types of sensory information from the environment, the integration and accumulation of sensory evidence in favour of a choice, and the motor instigation of that choice. Despite considerable progress in characterising unisensory perceptual decision making, relatively little is known about how this process changes when multisensory information must be integrated as a part of that decision. Specifically, it remains unclear how complementary auditory information alters the temporal evolution of neural activity during visual decision making. Here, we employ a modified paradigm whereby participants discriminated face and car images in a speeded task, and presented simultaneous, complementary sound stimuli during half of the trials. This modification allowed us to capture the temporal characteristics of any enhancement to electrophysiological signals during audiovisual decision making, compared to visual-only (unisensory) decision making. Discrimination increased gradually and peaked before participants made their choice, in a pattern mirroring that of evidence accumulation. When participants heard complementary sound stimuli, their decisions were more accurate, slower, and were best predicted by late post-sensory decision processing. Neurally-informed cognitive modelling further suggested that the addition of auditory information enhanced the rate of evidence accumulation during this period. Correspondingly, spatial representations of activity accompanying these changes featured a prominent centroparietal cluster.

2.2 Introduction

In everyday life, we often encounter situations that demand rapid decisions based on ambiguous sensory information. Consolidating the available evidence requires processing information presented in more than one sensory modality, a process commonly referred to as multisensory decision making (Angelaki, Gu and DeAngelis, 2009; Bizley, Jones and Town, 2016). For example, the decision to cross a street on a foggy morning will be based on a combination of visual evidence about hazy objects in your field of view and muffled sounds from various sources.

The presence of complementary audiovisual (AV) information can improve our ability to make perceptual decisions when compared to visual information alone (Lippert, Logothetis and Kayser, 2007; Raposo *et al.*, 2012; Kayser, Philiastides and Kayser, 2017). While recent studies have provided a detailed picture of the emergence of different types of unisensory and multisensory representations in the brain (Aller and Noppeney, 2019; Cao *et al.*, 2019; Rohe, Ehrlis and Noppeney, 2019), these studies have not provided a conclusive mechanistic account of how the brain encodes and ultimately translates the relevant sensory evidence into a decision (Bizley, Jones and Town, 2016). Specifically, it remains unclear whether the perceptual improvements by multisensory integration are best explained by a sensory processing benefit, changes in decision dynamics such as the efficiency of evidence accumulation, or the amount of accumulated evidence required to commitment to a choice.

These questions can be addressed within the general framework of sequential sampling models, such as the drift diffusion model, which posits that decisions are formed by a stochastic accumulation of evidence over time (Ratcliff, 1978; Ratcliff and Tuerlinckx, 2002; Ratcliff and Smith, 2004; Bogacz, 2007; Ratcliff and McKoon, 2008). The DDM decomposes behavioural data into internal processes that reflect the rate of the sensory accumulation process (drift rate), the amount of evidence required to make a decision (starting point and decision boundaries corresponding to the different decision alternatives), and latencies induced by early stimulus encoding and response production (non-decision time). Importantly, different signatures of brain activity were shown to reflect distinct aspects of this mechanistic model, and therefore, single trial measurements of

such neural activity could be used to further constrain these models (O’Connell, Dockree and Kelly, 2012; Philiastides, Heekeren and Sajda, 2014; Polanía *et al.*, 2014; Tagliabue *et al.*, 2019; von Lutz, Herding and Blankenburg, 2019).

To date, few studies have exploited such neural markers of dissociable representations associated with sensory and decision evidence to arbitrate between different accounts of how multisensory evidence influences decisions in the human brain. While some studies have performed careful comparisons between diffusion models and behavioural data (Drugowitsch *et al.*, 2014, 2015; Regenbogen *et al.*, 2016; Chandrasekaran, 2017; Colonius and Diederich, 2018), they did not constrain these models against neural activity. Other studies, in contrast, tried to dissociate pre- and post-perceptual mechanisms by traditional activation mapping but without clear mechanistic decision-making model supporting the interpretation of brain activity (Giard and Peronnet, 1999; Noppeney, Ostwald and Werner, 2010; Chandrasekaran, Lemus and Ghazanfar, 2013). Furthermore, many studies focusing on visual judgements have considered only very simplistic stimuli such as contrast, random dot motion, or orientation (Lippert, Logothetis and Kayser, 2007; Esposito, Mulert and Goebel, 2009; Hirokawa *et al.*, 2011; Leo *et al.*, 2011), which may be encoded locally at the level of early sensory processing, and hence may not generalise to complex real-life conditions. As a result, the general mechanistic influence of information from one modality on the decision making process of another modality remains unknown.

Here we employed a well-established visual object categorisation task in which early sensory evidence and post-sensory decision evidence can be properly dissociated based on EEG recordings. Specifically, using a face/car categorisation task, we have previously profiled two temporally distinct neural components that discriminate between the two stimulus categories; an ‘Early’ component, appearing approximately 170-200ms post-stimulus, and a ‘Late’ component, seen 300ms after the stimulus presentation (Delis *et al.*, 2016; Diaz *et al.*, 2017; Philiastides *et al.*, 2006; Philiastides and Sajda, 2006; Philiastides and Sajda, 2006; Philiastides and Sajda, 2007; Ratcliff *et al.*, 2009). We found that the Late component was a better predictor of behaviour than the Early component, predicted changes in the rate of evidence accumulation in a drift

diffusion model and shifted later in time with longer deliberation times. Taken together, these findings established that the Early component encodes the initial sensory evidence while the Late component encodes post-sensory decision evidence.

Finally, we capitalised on these distinct validated neural representations of visual information to identify the stage at which complementary auditory information influences the encoding of decision-relevant visual evidence in a multisensory context. Based on recent results (Rohe and Noppeney, 2015; Aller and Noppeney, 2019; Cao *et al.*, 2019), we hypothesised that using AV information to discriminate complex object categories, rather than more primitive visual features, would lead primarily to enhancements in the Late, as opposed to the Early, component, consistent with a post-sensory account. Importantly, by combining single trial modelling and EEG data, we exploited the trial-by-trial variability in the strength of the Early and Late neural components in a neurally-informed drift diffusion model to derive mechanistic insights into the specific role of these representations in AV integration.

2.3 Methods

2.3.1 Participants

We tested 40 participants (male = 18, female = 22; Mean age = 23.85, SD = 5.47) on a speeded (face versus car) categorisation tasks. All participants were right-handed with normal or corrected-to-normal vision and no self-reported history of neurological disorders. This study was approved by the ethics committee of the College of Science and Engineering at the University of Glasgow (CSE 300150102). All participants provided written informed consent prior to participation.

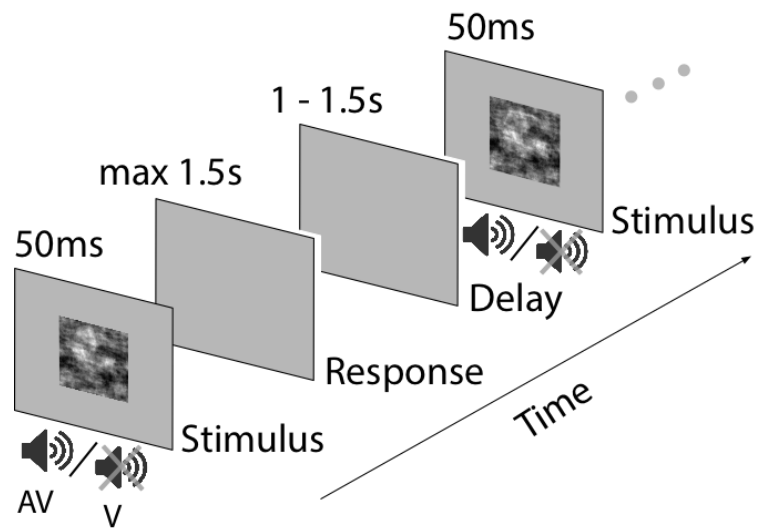


Figure 2.1 Experimental paradigm. Schematic representation of the task design illustrating the order of presented events on the testing day. Participants had to categorise noisy representations of faces and cars. A brief stimulus, which was either an image (V) or a congruent image and sound (AV), was presented for 50 ms and followed by a delay period of up to 1500 ms during which participants were required to indicate their decision with a button press. Their response was followed by an inter-trial interval (blank screen), jittered between 1000 and 1500 ms in duration, before the next stimulus was presented.

2.3.2 Stimuli and task

We used a set of 15 face and 15 car greyscale images (image size 670x670 pixels, 8-bits per pixel), adapted from our previous experiments (Philiastides and Sajda, 2006b; Philiastides, Ratcliff and Sajda, 2006; Diaz, Queirazza and Philiastides, 2017). Face images were selected from the face database of the Max Planck Institute of Biological Cybernetics (Troje and Bülthoff, 1996) and car images were sourced from the internet. Both image categories contained an equal number of frontal and side views (up to ± 45 degrees). All images were equated for spatial frequency, contrast, and luminance, and had identical magnitude spectra (average magnitude spectrum of all images in the database). We manipulated the phase spectra of the images using the weighted mean phase technique (Dakin *et al.*, 2002), whereby we changed the amount of visual evidence in the stimuli as characterised by their percentage phase coherence. To manipulate task difficulty, we used four levels of sensory visual evidence (27.5%, 30%, 32.5% and 35% phase coherence; for examples see Figure 2.1). These levels were based on our previous studies (Philiastides and Sajda, 2006b; Philiastides, Ratcliff and Sajda, 2006; Philiastides, Heekeren and Sajda, 2014; Diaz, Queirazza and Philiastides, 2017) as they are known to result in performance spanning psychophysical threshold. We displayed all pictures on light grey background (RGB [128, 128, 128]) using the PsychoPy software (version 1.83.04; Peirce, 2009).

Auditory sounds were used in addition to the visually presented images on a random half of trials. Sounds were either human speech or car/street-related sounds obtained from online sources. They were sampled at a rate of 22.05 kHz and stored as .wav files. In MATLAB (version 2015a, The MathWorks, 2015), we added a 10 ms cosine on/off ramp to reduce the effects of sudden sound onsets and normalised all sounds. Subsequently, we reduced the intensity of these normalised sounds by lowering their amplitude by 80%. Sounds were embedded in Gaussian white noise and the relative amplitude of the sounds and noise was manipulated to create 17 different levels of relative noise-to-signal ratios (ranging from 12.5% to 200% of noise relative to the lowered amplitude signal, in increments of 12.5%). The resulting noisy speech and car-related sounds were presented binaurally for 50 ms through Sennheiser stereo headphones HD 215.

The stimulus display was controlled by a Dell 64 bit-based machine (16 GB RAM) with an NVIDIA Quadro K620 (Santa Clara, CA) graphics card running Windows Professional 7 or Linux-x86_64 and PsychoPy presentation software (version 1.83.04; Peirce, 2009). All images were presented on an Asus ROG Swift PG278Q monitor (resolution, 2560x1440 pixels; native refresh rate 144 Hz, set to 120 Hz). Participants were seated 75 cm from the stimulus display, and each image subtended approximately 11 x 11 degrees of visual angle.

Task. We employed an adapted audiovisual version of the widely used visual face versus car image categorisation task (Philiastides and Sajda, 2006b, 2007; Philiastides, Ratcliff and Sajda, 2006; Diaz, Queirazza and Philiastides, 2017). This task required participants to decide whether they saw a face or a car embedded in the stimulus (Figure 2.1). Participants were asked to indicate their decision via button press on a standard keyboard as soon as they had formed a decision. The response deadline was set at 1.5 seconds. On half of the trials, participants were also given an additional auditory cue in the form of a brief noisy sound that was congruent with the picture's content. Audiovisual face trials were accompanied by a human speech sound, whereas audiovisual car trials were accompanied by a car-related sound, such as squeaking tires or a slammed door. All stimuli were presented for 50 ms in the centre of the screen, and on audiovisual trials to both ears. During audiovisual trials pictures and sounds were presented simultaneously. More specifically, we used four levels of visual noise, but only one (participant-specific) auditory difficulty level, obtained at perithreshold performance during an initial auditory training task (see below). Thereby we accounted for inter-individual differences in auditory perception, independently of visual image difficulty.

Training. This experimental paradigm required participants to attend a training and a testing session on two consecutive days, at the same time of the day. On the first day (i.e., the training day), participants were asked to perform three separate simple categorisation tasks to familiarise themselves with the task: (1) a visual image discrimination task (face versus car), (2) an auditory sound discrimination task (face/speech versus car/street sounds) and (3) an audiovisual discrimination task (face versus car). Only during training, participants were given visual feedback following each response on all three tasks in the form of

visual feedback presented in the centre of the screen (“Incorrect” written in red and “Correct” written in green for trials on which participants responded within the response deadline, and “Too slow” written in blue when they exceeded the response deadline).

During the visual training task, we used the same images and all four levels of visual evidence as on the second day (i.e., the testing day). During the auditory training task, we presented sounds to participants using eight different levels of relative noise-to-signal ratios (12.5%, 37.5%, 62.5%, 93.75%, 125%, 150%, 175%, and 200% of added noise). We estimated subject-specific noise levels supporting individual perithreshold performance (i.e., ~70% decision accuracy), including levels that might have fallen in between the eight noise-to-signal ratios used in this training task (from the larger set of 17; $M = 140\%$, $SD = 45\%$). We used these individual levels for the audiovisual training task and the main experiment. During the audiovisual training task, we used all images at the four levels of visual evidence together with the subject-specific perithreshold noise-level determined above. This audiovisual training task mimicked the main task presented on the second (testing) day, with the exception that participants received feedback on their choices.

Overall, on the training day, we presented 480 trials for each of the visual and auditory discrimination training tasks split into four blocks of 120 trials with a 60 second rest period between blocks. We presented 240 trials, split into two blocks, during the audiovisual training task. Taken together, all three training tasks lasted approximately 55 minutes on the first day.

Full task. On the second day, we collected behavioural and EEG data using randomly interleaved visual (unisensory) and audiovisual (multisensory) trials in a combined task (Figure 2.1). Stimuli presentation employed the same task timings as outlined above on both days. Crucially, we did not provide any feedback to participants during testing. Using only one auditory noise level per participant on the testing day allowed us to evaluate the effects of auditory benefit at different levels of visual evidence. We presented 720 trials, divided equally between all stimulus categories (i.e., face/car, visual/audiovisual, and four levels of visual evidence), in short blocks of 60 trials with 60 second breaks

between blocks. The entire task on the testing day lasted approximately 45 minutes. EEG data were collected only during the testing day.

2.3.3 Behavioural analysis

Our main behavioural analysis quantified participants' behavioural performance (i.e., decision accuracy and response times) in the data collected during the testing day using two separate generalised linear mixed effects models (GLMMs). GLMMs are superior to traditional repeated measures ANOVA analysis as their random effects structure better accounts for inter-participant variability and allows mixing of categorical and continuous variables (Baayen, Davidson and Bates, 2008). Both models included all main effects and interactions of our two predictor variables, modality (visual and audiovisual) and visual evidence (27.5%, 30%, 32.5% and 35%), along with by-subject random slopes and random intercepts for all relevant main effects. Hence, the two models used the maximal random effects structure justified by the design (Barr *et al.*, 2013). We employed post-hoc likelihood-ratio X^2 model comparisons to quantify the predictive power and significance of all main effects and interactions initially showing p values below or around threshold (i.e., $\alpha < 0.05$) by both GLMMs. These likelihood-ratio X^2 model comparisons compared the full model (i.e., a model including all main effects, interactions and random effects) to a reduced model excluding the predictor or the set of predictors in question. Only results and statistics of the post-hoc model comparisons are reported in the main results section. We performed these GLMM analyses using the *lme4* package (Bates *et al.*, 2015) in RStudio (RStudio Team, 2016) specifying a *binomial logit* model in the family argument of the *glmer* function for decision accuracy, a binary dependent variable, and a *gamma* model for response time, a continuous dependent variable while selecting the *bobyqa* optimiser. The predictor modality was entered in mean-centred form (deviation coding), whereas the predictor visual evidence (four levels) was entered using mean-centred backward difference coding. By using mean-centred coding schemes we accounted for small imbalances in trial numbers between the predictor's levels. Random correlations were excluded for both GLMMs.

To formally rule out that our choice of subject-specific levels of auditory evidence could exclusively explain individual improvements in decision accuracy on audiovisual trials, we correlated these measures across subjects using a robust bend correlation analysis (Pernet, Wilcox and Rousselet, 2013). Specifically, we evaluated whether the individual levels of auditory noise correlated with the difference in accuracy between visual and audiovisual trials (i.e., $\text{accuracy}_{\text{audiovisual}} - \text{accuracy}_{\text{visual}}$) across participants. As part of this correlation analysis, we computed the mean accuracy across all trials of each level of visual evidence and modality for each participant separately. We found that the level of subject-specific auditory noise accounted for only a minimal fraction of the variance in accuracy improvements ($R^2 = 0.01$).

2.3.4 EEG data acquisition and pre-processing

We acquired continuous EEG data in a sound-attenuated and electrostatically shielded room from a 64-channel EEG amplifier system (BrainAmps MR-Plus, Brain Products, Germany) with Ag/AgCl scalp electrodes placed according to the international 10-20 system on an EasyCap (Brain Products GmbH, Germany). A chin electrode acted as ground and all channels were referenced to the left mastoid during recording. We adjusted the input impedance of all channels to $<20\text{k}\Omega$. The data were sampled at a rate of 1000 Hz and underwent online (hardware) filtering by a 0.0016-250 Hz analogue band-pass filter. We used PsychoPy and Brain Vision Recorder (BVR; Version 1.10, Brain Products, Germany) to record trial specific information including experimental event codes and button responses simultaneously with the EEG data. These data were collected and stored for offline analysis in MATLAB. Offline data pre-processing included applying a software-based 0.5-40 Hz band-pass filter. To avoid phase-related distortions, we applied these filters non-casually (using MATLAB “*filtfilt*”). Finally, the EEG data were re-referenced to the average of all channels.

We removed eye movement artefacts such as blinks and saccades using data from an eye movement calibration task completed by participants before the main task on the testing day. During this task, participants were instructed to

blink repeatedly upon the appearance of a black fixation cross on light grey background in the centre of the screen before making several lateral and horizontal saccades according to the location of the fixation cross on the screen. Using principal component analysis, we identified linear EEG sensor weights associated with three eye movement artefacts (one component each for upward/downward saccades, leftward/rightward saccades, and blinks), which were then projected onto the broadband data from the main task and subtracted out (Parra, Spence, Gerson, and Sajda, 2005). The choice of principal component analysis to identify and remove these artefacts was based on previous experiments in our lab having successfully done so. Further, using independent component analysis in its place would have led to almost identical results as the topographies of these components are so prototypical. We excluded all trials from all subsequent analyses where participants exceeded the response time limit of 1.5 seconds, indicated a response within less than 300 ms after onset of the stimulus or the EEG signal exceeded a maximum amplitude of 150 μV during the trial (0.8%, 0.06%, and 0.03% of all trials across participants, respectively). As we excluded trials where participants failed to respond in time (i.e., within 1.5 seconds), this also meant that any trials where participants blinked and missed the stimulus (as it was presented for only 50 ms) should also have been excluded.

2.3.5 EEG data analysis

We employed a linear multivariate single-trial discriminant analysis of stimulus- and response-locked EEG data (Parra *et al.*, 2002; Parra *et al.*, 2005) to identify early sensory and late decision-related EEG components discriminating between face and car trials as in previous work (Philiastides and Sajda, 2006b; Ratcliff, Philiastides and Sajda, 2009). We performed this analysis separately for visual and audiovisual trials to independently identify the sensor signals discriminating the relevant visual evidence in each modality and allow direct comparisons between them in terms of overall discrimination performance.

Specifically, we identified a projection of the multichannel EEG signal, $x_i(t)$, where $i = [1 \dots N \text{ trials}]$, within short time windows (i.e., a sliding window

approach) that maximally discriminated between face and car trials (i.e., visual discrimination: face vs car; audiovisual discrimination: face/speech vs car/street sounds). All time windows had a width of 60 ms and onset intervals every 10 ms. These windows were centred on and shifted from -100 to 1,000 ms relative to stimulus onset on stimulus-locked data and from -600 to 500 ms relative to the response button press on response-locked data. Specifically, a 64-channel spatial weighting $\mathbf{w}(\tau)$ was learned by means of logistic regression (Parra *et al.*, 2005) that achieved maximal discrimination within each time window, arriving at the one-dimensional projection $y_i(\tau)$, for each trial i and a given window τ :

$$\mathbf{y}(\tau) = \mathbf{w}^T \mathbf{x}(\tau) = \sum_{i=1}^D w_i x_i(\tau) \quad (1)$$

Here, T refers to the transpose operator. In separating the two stimulus categories, the discriminator was designed to map component amplitudes $y_i(\tau)$ for face and car trials, to positive and negative values, respectively. To quantify the performance of our discriminator for each time window, we used the area under a receiver operating characteristic (ROC) curve (Green and Swets, 1966), referred to as an A_z value, combined with a leave-one-trial-out cross-validation procedure to control for overfitting (Duda, Hart and Stork, 2001; Philiastides and Sajda, 2006a; Gherman and Philiastides, 2018).

Specifically, for every iteration, we used $N-1$ trials to estimate a spatial filter \mathbf{w} , which was then applied to the left-out trial to obtain out-of-sample discriminant component amplitudes (\mathbf{y}) and compute the A_z value. Moreover, we determined significance thresholds for the discriminator performance (rather than assuming an A_z of 0.5 as chance performance) using a bootstrap analysis whereby face and car labels were randomised and submitted to a separate leave-one-trial-out test. This randomisation procedure was repeated 1000 times, producing a probability distribution for A_z , which we used as reference to estimate the A_z value leading to a significance level of $P < 0.05$ (subject average $A_z \text{sig} = 0.56$). Note that this EEG analysis pipeline was performed on individual subjects such that each subject became their own replication unit (Smith and Little, 2018).

Finally, the linearity of our model allowed us to compute scalp projections of our discriminating components resulting from equation (1) by estimating a forward model as:

$$\mathbf{a}(\tau) = \frac{\mathbf{x}(\tau)\mathbf{y}(\tau)}{\mathbf{y}(\tau)^T\mathbf{y}(\tau)} \quad (2)$$

where the EEG data (\mathbf{x}) and discriminating components (\mathbf{y}) are now in a matrix and vector notation, respectively, for convenience. Such forward models can be displayed as scalp topographies and interpreted as the coupling between the observed EEG and the discriminating component amplitudes (i.e., vector \mathbf{a} reflects the electrical coupling of the discriminating component \mathbf{y} that explains most of the activity in \mathbf{x}).

2.3.5.1 Temporal cluster-based bootstrap analysis

To quantify if and when the discriminator performance differed between visual and audiovisual trials, we used a percentile bootstrap technique for comparing the group-level A_z difference between two dependent samples (Rousselet, Foxe and Bolam, 2016; Rousselet, Pernet and Wilcox, 2017). Specifically, on a sample-by-sample basis, we created a distribution of shuffled A_z difference scores (i.e., audiovisual minus visual) across participants (drawing with replacement). We repeated this shuffling procedure 1000 times for each sample whereby we created a random bootstrap distribution of median A_z difference scores from every iteration. We computed the median of this bootstrap distribution for a given sample along with the 90% confidence interval (5% to 95%) of the resulting distribution of median difference scores. To test whether our bootstrapped median difference was significantly different from zero for each sample we compared it against the lower bound of the estimated confidence interval (i.e. at the 5%; $P < 0.05$).

To form contiguous temporal clusters and avoid transient effects due to false positives, we required a minimum temporal cluster size of at least three significant samples. This threshold was determined by means of the 95th percentile of a data-driven null distribution of maximum cluster sizes. We first applied a permutation procedure (i.e., shuffling temporal samples without replacement) to abolish the relationship across temporal samples, while keeping the relative difference between V and AV A_z values unchanged, for each sample and participant. We generated the null distribution of maximum cluster sizes by

calculating the maximum number of adjacent significant samples of the largest cluster for each of the 1000 iterations. This procedure corrects for multiple comparisons and is comparable to the temporal cluster-based non-parametric permutation test reported in Maris and Oostenveld (2007). We performed this analysis on the discriminator performance on both stimulus- and response-locked data (Figure 2.3b and Figure 2.4b respectively).

Lastly, to ensure that neural effects were also reliably traceable on individual participants without group-level averages masking variability, we also computed the proportion of participants who demonstrated a participant-level effect in line with the general group-level effect per sample (that is, higher audiovisual A_z value for a given sample - see Figure 2.3c and Figure 2.4c). We performed these statistical analyses using MATLAB code obtained from the Figshare and Github repositories associated with Rousselet *et al.* (2017) and Rousselet *et al.* (2016).

2.3.5.2 Hierarchical Drift Diffusion Modelling of behavioural data

We fit the subjects' performance, i.e. face/car choice and response time (RT), with a hierarchical drift diffusion model (HDDM) (Wabersich and Vandekerckhove, 2014). Similar to the traditional drift diffusion model, the HDDM assumes a stochastic accumulation of sensory evidence over time, toward one of two decision boundaries representing the two choices (face or car). The model returns estimates of internal components of processing such as the rate of evidence accumulation (drift rate), the distance between decision boundaries controlling the amount of evidence required for a decision (decision boundary), a possible bias towards one of the two choices (starting point) and the duration of non-decision processes (non-decision time), which include stimulus encoding and response production.

HDDM model fitting. The HDDM uses Markov-chain Monte Carlo sampling to iteratively adjust the above parameters to maximize the summed log likelihood of the predicted mean response time (RT) and accuracy. The DDM parameters were estimated in a hierarchical Bayesian framework, in which prior distributions of the model parameters were updated on the basis of the

likelihood of the data given the model, to yield posterior distributions (Kruschke, 2010a; Wiecki, Sofer and Frank, 2013; Wabersich and Vandekerckhove, 2014). The use of Bayesian analysis, and specifically the hierarchical drift diffusion model, has several benefits relative to traditional DDM analysis. First and foremost, this framework supports the use of other variables as regressors of the model parameters to assess relations of the parameters with other physiological or behavioural data (Cavanagh *et al.*, 2014; Frank *et al.*, 2015; Nunez, Srinivasan and Vandekerckhove, 2015; Turner, Van Maanen and Forstmann, 2015; Nunez, Vandekerckhove and Srinivasan, 2017; Pedersen, Frank and Biele, 2017; Delis *et al.*, 2018). This property of the HDDM allowed us to establish the link between the EEG components and the aspects of the decision making process they are implicated in. Second, posterior distributions directly convey the uncertainty associated with parameter estimates (Kruschke, 2010b). Third, the Bayesian hierarchical framework has been shown to be especially effective when the number of observations is low (Ratcliff and Childers, 2015). Fourth, within this hierarchical framework, all observers in a dataset are assumed to be drawn from a group, which yields more stable parameter estimates for individual subjects (Wiecki, Sofer and Frank, 2013).

To implement the hierarchical DDM, we used the Wiener module (Wabersich and Vandekerckhove, 2014) in JAGS (Plummer, 2003), via the Matjags interface in MATLAB to estimate posterior distributions. For each trial, the likelihood of accuracy and RT was assessed by providing the Wiener first-passage time (WFPT) distribution with the three model parameters (boundary separation, non-decision time, and drift rate). Parameters were drawn from uniformly distributed priors and were estimated with non-informative mean and standard deviation group priors. The starting point was set as the midpoint between the two decision boundaries as the experimental design induced no bias towards one of the two choices (Philiastides *et al.*, 2011; Herz *et al.*, 2016). For each model, we ran 5 separate Markov chains with 5500 samples each; the first 500 were discarded (as “burn-in”) and the rest were subsampled (“thinned”) by a factor of 50 following the conventional approach to MCMC sampling whereby initial samples are likely to be unreliable due to the selection of a random starting point and neighbouring samples are likely to be highly correlated (Wiecki, Sofer and Frank, 2013; Wabersich and Vandekerckhove, 2014). The remaining samples constituted the

probability distributions of each estimated parameter from which individual parameter estimates were computed.

To ensure convergence of the chains, we computed the Gelman-Rubin R2 statistic (which compares within-chain and between-chain variance) and verified that all group-level parameters had an R2 close to 1 and always lower than 1.03. For comparison between models, we used the Deviance Information Criterion (DIC), a measure widely used for fit assessment and comparison of hierarchical models (Spiegelhalter, Best, Carlin, & Van Der Linde, 2002). DIC selects the model that achieves the best trade-off between goodness-of-fit and model complexity. Lower DIC values favour models with the highest likelihood and least degrees of freedom.

HDDM model with EEG regressors. We first estimated a model that used our EEG discrimination analysis to inform the fitting of the behavioural data. In this model, we input the single-trial RTs and (face or car) choices of all 40 subjects and hypothesized that the evidence accumulation rate on each trial would be dependent on the amount of neural evidence about face or car choice on that trial. Therefore, as part of the model fitting within the HDDM framework, we used the single-trial EEG measures of face.vs.car discrimination as regressors of the drift rate (δ) as follows:

$$\delta = \beta_0 + \beta_1 * y_{Early}^S + \beta_2 * y_{Late}^R \quad (3)$$

where y_{early}^S and y_{late}^R are the single-trial discriminator amplitudes of subject-specific stimulus-locked Early EEG components (corresponding to individual peak A_z across V and AV in the range 170-250ms post-stimulus) and response-locked Late EEG components (corresponding to individual peak A_z difference between AV and V in the range -150ms to -60ms pre-response), respectively. Whereas the analysis used to identify individual peak A_z was implemented to understand the temporal characteristics of audiovisual decision making, and therefore to test our hypothesis that the Late component would capture the effects of the additional sensory evidence provided, our analysis using single-trial measures as explained here instead sought to link neural activity with behaviour and understand the functional role of the component that produced those values. Had the same hypothesis been tested by both stages of analysis, the results

would have been inflated, however here each stage is motivated by a distinct hypothesis, specifically the temporal characteristics and then the link between these EEG components and behavioural performance.

The coefficients β_i weight the slope of the drift rate by the values of y_{early}^S and y_{late}^R on that specific trial, with an intercept β_0 . Here we estimated β_i 's for each subject, sensory condition and coherence level. Hence, by using these regression coefficients we were able to test the influences of each of the two identified components on the drift rate in both sensory conditions (Cavanagh *et al.*, 2014). Posterior probability densities of each regression coefficient were estimated using the sampling procedure described above. Significantly positive (negative) effects were determined when >99.9% of the posterior density was higher (lower) than 0.

HDDM model without neural information. For comparison, we also estimated a HDDM without including any neural correlates. We fit the HDDM to RT distributions for correct and incorrect choices conditioned on the sensory condition (V or AV) for each trial. Overall drift rate, boundary separation and non-decision time were estimated for each individual and were dependent on the sensory condition. As per common practice, we assumed that evidence strength affected the drift rate, thus we separately estimated drift rate for each coherence level of the two sensory conditions (Ratcliff and Frank, 2012).

2.4 Results

2.4.1 Behavioural results

We collected behavioural and EEG data from 40 participants during a speeded face/car categorisation task (Figure 2.1). Participants were required to identify a noisy image as being either a face or a car, presented in a randomly interleaved fashion either alone (visual trials; V) or simultaneously with distorted speech or car sounds (audiovisual trials; AV). The amount of visual evidence (image phase coherence) varied consistently across subjects over four levels, whereas auditory evidence was set at a subject-specific threshold throughout the task. This was determined by calculating the level of distortion required for between 68% and 72% of trials being correctly discriminated during an auditory-only training session on the previous day (as previously described in ‘Training’, page 26).

We used generalised linear mixed-effects models (GLMM) and post-hoc likelihood-ratio χ^2 model comparisons to evaluate decision accuracy and response times (using a binomial logit and a gamma model, respectively), both as a function of modality (V/AV) and the levels of visual evidence (see **Methods**). We found that participants performed more accurately on trials with AV evidence ($\chi^2 = 30.04$, $df = 1$, $P < 0.001$) as well as with increases in the amount of visual evidence ($\chi^2 = 87.13$, $df = 3$, $P < 0.001$) (Figure 2.2a,b). There were no significant interactions between modality and the level of visual evidence (all $P > 0.05$), though, on average, the AV accuracy improvements appear most enhanced at the lowest (most difficult) level of visual evidence. Response times increased with AV evidence ($\chi^2 = 18.73$, $df = 1$, $P < 0.001$) and decreased with the amount of visual evidence ($\chi^2 = 16.03$, $df = 3$, $P = 0.001$) (Figure 2.2c,d). There were no significant interactions between modality and the level of visual evidence (all $P > 0.05$).

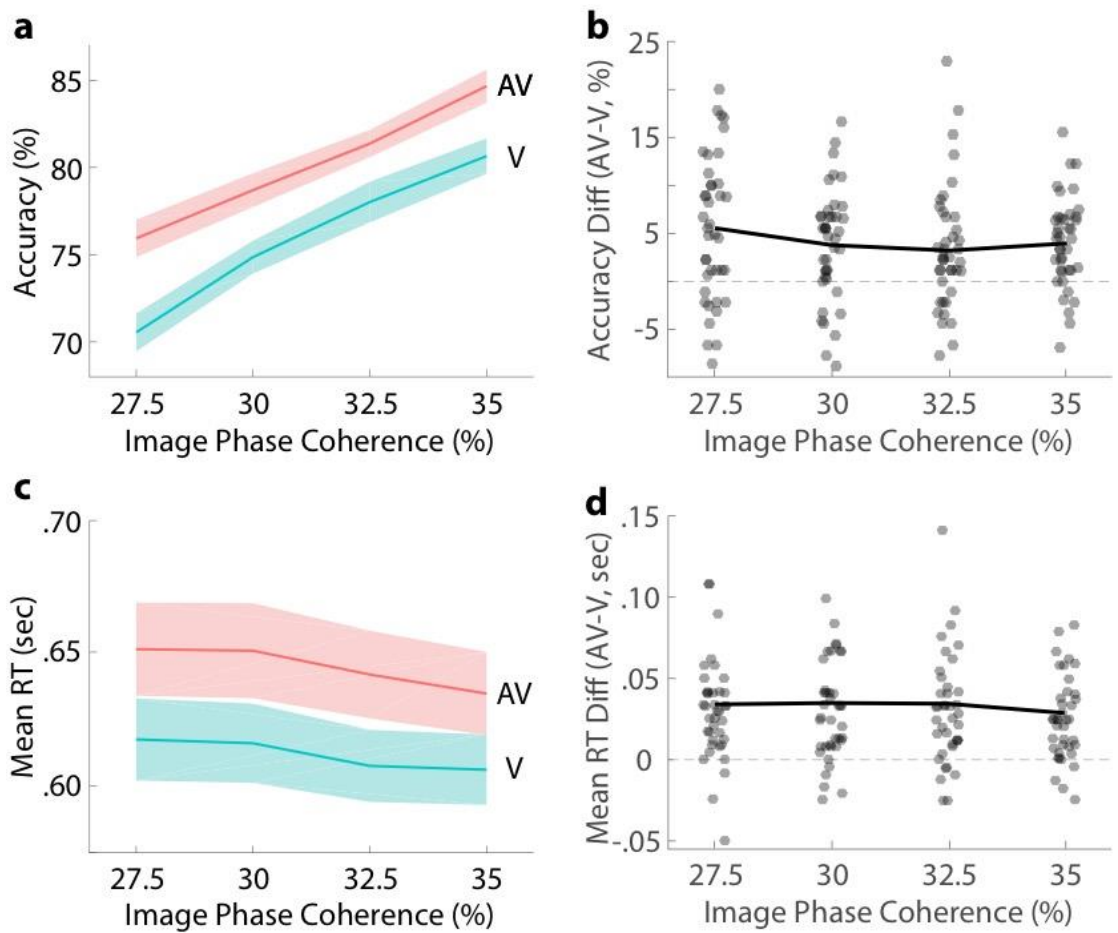


Figure 2.2 Behavioural performance. **A,C**, Group averages of (a) decision accuracy and (c) response time across the four levels of visual evidence (phase coherence) and as a function of the visual (V; turquoise) and audiovisual (AV; red) trials. Shaded error bars indicate standard errors across participants (N=40). **B,D**, Individual participant behavioural performance changes (audiovisual - visual trials) for (b) decision accuracy and (d) response time across the four levels of visual evidence (phase coherence). Solid black lines indicate group averages.

To ensure that our choice in the amount of subject-specific auditory evidence could not independently explain the overall improvements in accuracy during AV trials, we quantified the extent to which subjects provided with more accurate auditory evidence benefited more in AV trials. We found that the amount of auditory evidence explained only a minimal fraction of the variance in accuracy across subjects ($R^2 = 0.01$). Taken together, these results suggest that the combined influence of audiovisual information indeed contributed to an increased likelihood of making a correct decision (overall improvement $M = 4.14\%$, $SD = 3.91$), but at the cost of some speed (overall slowing $M = 33.1\text{ms}$, $SD = 35.02$). The latter is likely due to additional encoding time required for the auditory stimulus (see **Neurally-informed cognitive modelling**).

2.4.2 Temporal impact of auditory evidence on visual representations

Next, we analysed the EEG data to identify the Early (sensory) and Late (decision-related) components that discriminated between face and car visual evidence. We performed this analysis separately for V and AV trials to characterise the extent to which the visual representations encoded in these temporally distinct components were affected by the additional auditory evidence. Specifically, for each subject separately, we performed a single-trial multivariate discriminant analysis (Parra *et al.*, 2005; Sajda, Philiastides, & Parra, 2009) to estimate linear spatial weightings (i.e. spatial filters) that maximally discriminated face-vs-car trials within short pre-defined temporal windows, locked either to the onset of the stimulus or the response (see **Methods**).

Applying the resulting spatial filters to single-trial data produces a measure of the discriminating component amplitudes (henceforth y), which can be used as an index of the quality of the visual evidence in each trial (Philiastides and Sajda, 2006b; Philiastides, Ratcliff and Sajda, 2006; Guggenmos, Sterzer and Cichy, 2018). In other words, more extreme amplitudes, positive or negative, indicate more face or car evidence respectively, while values closer to zero indicate less evidence. To quantify the discriminator's performance over time and identify our Early and Late components, we used the area under a receiver operating characteristic curve (i.e., A_z value) with a leave-one-trial-out cross validation approach, to control for overfitting.

The discriminator's performance as a function of stimulus-locked time revealed the presence of two temporally specific components (Figure 2.3a; Early (V/AV): mean peak time 230/220 ms; Late (V/AV): mean peak time 460/500ms), with distinct scalp topographies (Figure 2.3a; Early: bilateral occipitotemporal clusters, consistent with the well-described N170; Late: centroparietal cluster, consistent with the decision-related centroparietal positivity) but similar spatial projections for V and AV trials (group average cross correlation: $\text{Early}_{V/AV} = 0.97$; $\text{Late}_{V/AV} = 0.91$). These components (and overall classification performance)

were comparable to those identified in previous experiments using our face/car paradigm, both in terms of their latency and spatial distribution (Philiastides and Sajda, 2006a, 2006b, 2007; Philiastides, Ratcliff and Sajda, 2006; Ratcliff, Philiastides and Sajda, 2009).

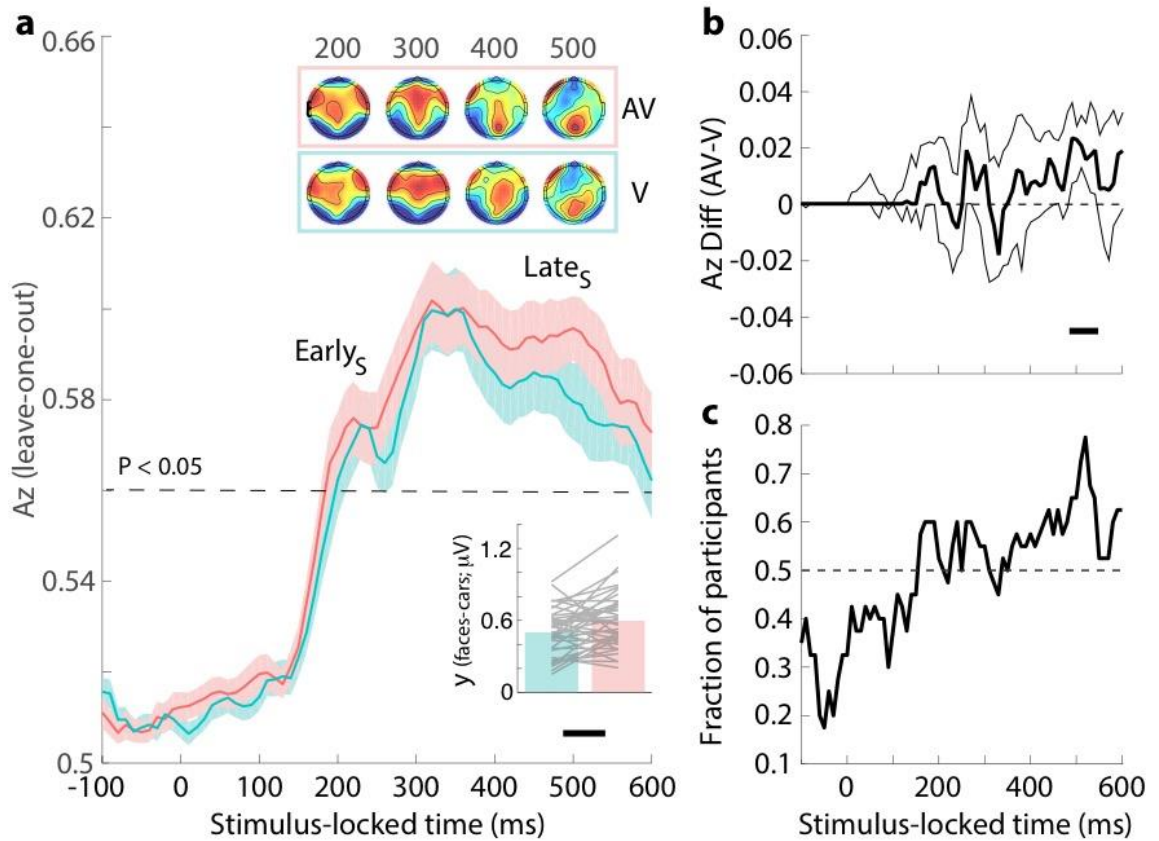


Figure 2.3 Stimulus-locked face-vs-car discrimination analysis. **a**, Mean discriminator performance (A_z) during face versus car discrimination of stimulus-locked EEG data after a leave-one-trial-out cross-validation procedure, as a function of the visual (V; turquoise) and audiovisual (AV; red) conditions. Dotted black line represents the group average permutation threshold at $P < 0.05$. Shaded error bars indicate standard errors across subjects. Scalp topographies at representative time windows corresponding to the Early and Late EEG components, encoding sensory and post-sensory visual evidence, respectively. Inset: Late EEG component amplitudes reflecting the relative separation across face and car trials ($y_{\text{faces}} - y_{\text{cars}}$) at the point of maximum A_z separation between V and AV trials (solid black line – see **b**), **b**, Bootstrapped difference in discriminator performance (audiovisual - visual; thick black line) with 90% confidence intervals (5-95%; thin black lines). Horizontal thick black lines above the x-axes in panels **a** and **b** illustrate significant temporal windows resulting from this permutation testing (i.e. those in which the lower confidence interval is greater than zero with an added minimum requirement of three contiguous windows). **c**, Fraction of participants showing discriminator performance (A_z) in the same direction as the group-level mean.

Having identified these components in both V and AV trials we next used a temporal cluster-based permutation analysis (Rousselet, Pernet and Wilcox, 2017), to identify contiguous windows during which the discriminator performance differed systematically between V and AV trials. Specifically, for each temporal sample we created a bootstrap distribution of group-level A_z difference scores (AV - V) and compared our bootstrapped median difference score against the estimated confidence intervals of the distribution (supporting a significance level of $P < 0.05$). To form contiguous temporal clusters and avoid transient effects due to false positives, we required a minimum temporal cluster size of at least three significant samples (see **Methods**).

This analysis revealed only a single temporal cluster overlapping with the Late component (490ms to 540ms) over which the discriminator performance for AV trials was significantly improved compared to V trials (Figure 2.3a,b). During this time up to 78% of participants showed increases in the discriminator's performance for AV trials, compared to only 60% of participants during the Early component (Figure 2.3c). Taken together, these findings indicate that the addition of auditory information in our task enhances primarily the quality of visual evidence (as reflected in our discriminator component amplitudes y) during post-sensory decision-related processing of our face/car stimuli (Figure 2.3a; inset).

In previous work, we showed that the Late component activity starts out as being stimulus-locked but persists and becomes more robust near the response (Philiastides and Sajda, 2006a, 2006b; Philiastides, Ratcliff and Sajda, 2006; Blank *et al.*, 2013), consistent with the notion that decision evidence reverberates and accumulates continuously until one commits to a choice. We therefore repeated the single-trial multivariate discrimination analysis on response-locked data. Importantly, this analysis also helps rule out potential motor confounds associated with differences in response times across V and AV trials by abolishing potential temporal lags near the time of the response.

As with the stimulus-locked analysis, we ran a cluster-based permutation test, comparing face/car discriminator performance for V and AV trials. We identified a temporal cluster leading up to the eventual choice (-150ms to -60ms pre-response) during which discriminator performance was significantly enhanced for

AV compared to V trials (Figure 2.4a,b), with consistent effects (>70%) appearing across participants (Figure 2.4c). Inspection of the resulting scalp maps during this period indicated that the spatial topographies, featuring a prominent centroparietal cluster, are consistent with the Late component seen in the earlier stimulus-locked analysis (compare scalp topographies in Figure 2.3a/Figure 2.4a) and in line with previous work (Blank *et al.*, 2013). These findings further highlight that it is primarily late, decision-related visual evidence that is being amplified during audiovisual object categorisation (Figure 2.4a; inset).

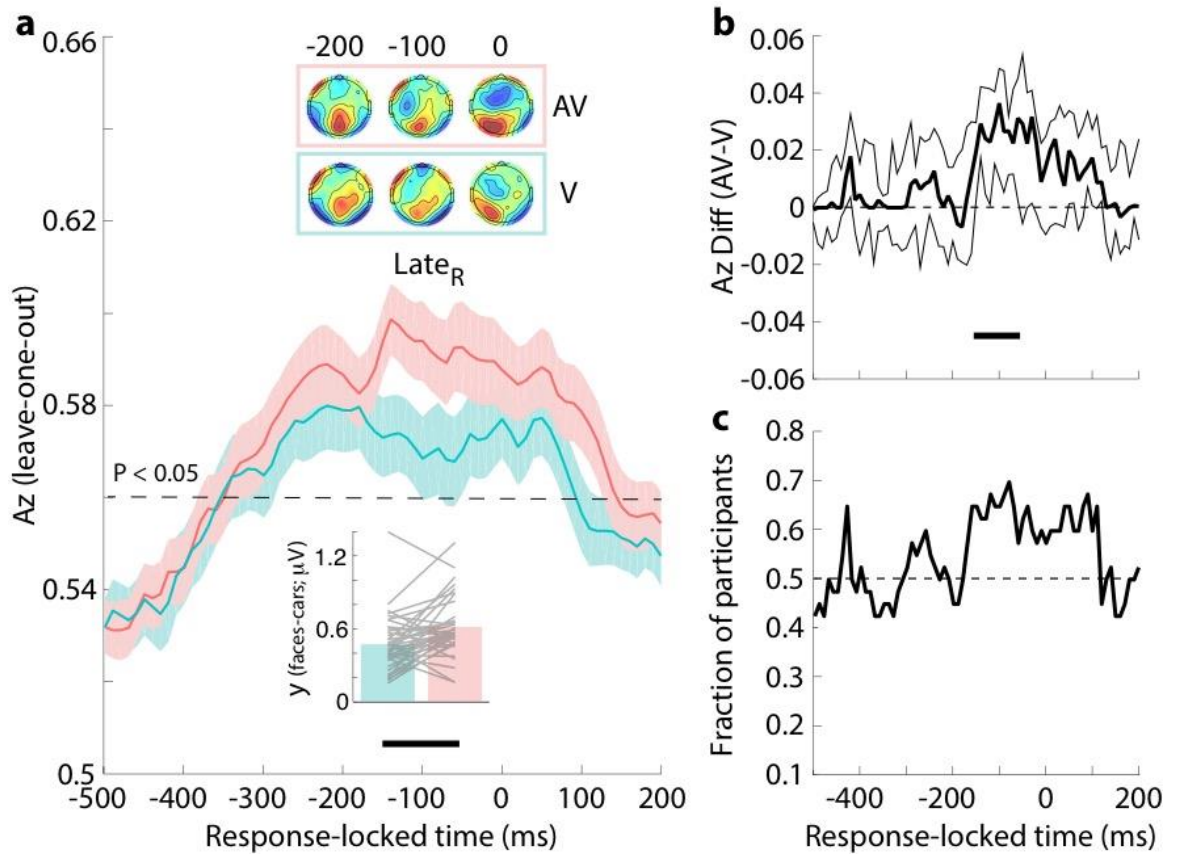


Figure 2.4 Response-locked face-vs-car discrimination analysis. **a**, Mean discriminator performance (A_z) during face versus car discrimination of response-locked EEG data after a leave-one-trial-out cross-validation procedure, as a function of the visual (V; turquoise) and audiovisual (AV; red) conditions. Dotted black line represents the group average permutation threshold at $P < 0.05$. Shaded error bars indicate standard errors across subjects. Scalp topographies at representative time windows corresponding to the Late EEG component, encoding persistent post-sensory visual evidence up until the eventual commitment to choice. Inset: Late EEG component amplitudes reflecting the relative separation across face and car trials ($y_{\text{faces}} - y_{\text{cars}}$) at the point of maximum A_z separation between V and AV trials (solid black line – see **b**). **b**, Bootstrapped difference in discriminator performance (audiovisual - visual; thick black line) with 90% confidence intervals (5-95%; thin black lines). Horizontal thick black lines above the x-axes in panels a and b illustrate significant temporal windows resulting from this permutation testing (i.e. those in which the lower confidence interval is greater than zero with an added minimum requirement of three contiguous windows). **c**, Fraction of participants showing discriminator performance (A_z) in the same direction as the group-level mean.

2.4.3 Neurally-informed cognitive modelling

Having characterised whether the added influence of auditory information enhances early sensory or late post-sensory visual representations, we then asked whether the identified single-trial neural responses are directly linked to improvements in behaviour between V and AV trials. To this end, we employed a neurally-informed variant of the traditional Hierarchical Drift Diffusion Model (HDDM, see **Methods**), a well-known psychological model for characterising rapid decision making (Ratcliff and McKoon, 2008; Wiecki, Sofer and Frank, 2013; Nunez, Vandekerckhove and Srinivasan, 2017) to offer a mechanistic account of how the human brain translates the relevant evidence into a decision. In doing so, we directly constrained the model based on additional neural evidence, hence closing this persistent gap in the literature (Noppeney, Ostwald and Werner, 2010; Chandrasekaran, Lemus and Ghazanfar, 2013; Drugowitsch *et al.*, 2014).

In brief, the traditional HDDM decomposes task performance (i.e. choice and RT), into internal components of processing representing the rate of evidence integration (drift rate, δ), the amount of evidence required to make a choice (decision boundary separation, α), and the duration of other processes, such as stimulus encoding and response production (non-decision time, τ). Ultimately, by comparing the obtained values of all three core HDDM parameters across the V and AV trials, we could associate any behavioural differences resulting from the addition of auditory information (improved performance and longer RTs as in Figure 2.1) to the constituent internal processes reflected by each of these parameters.

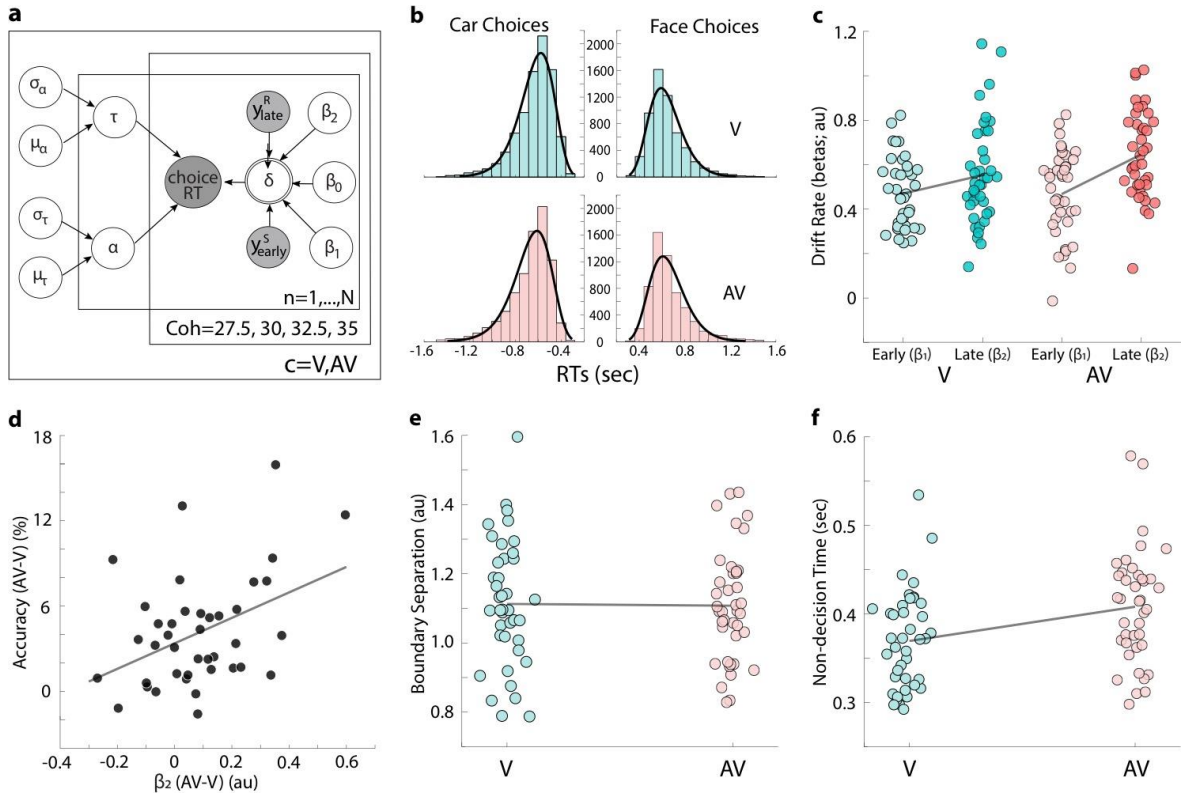


Figure 2.5 Neurally-informed cognitive modelling. **a**, Graphical representation showing hierarchical estimation of nHDDM parameters. Round nodes represent continuous random variables and double-bordered nodes represent deterministic variables, defined in terms of other variables. Shaded nodes represent recorded or computed signals, i.e. single-trial behavioural data (accuracy, RT) and EEG component amplitudes (y 's). Parameters are modelled as random variables with inferred means μ and variances σ^2 . Plates denote that multiple random variables share the same parents and children. The outer plate is over sensory conditions (V,AV) and the two inner plates are over phase coherence levels (Coh) and subjects (n) respectively. **b**, Histogram and nHDDM model fits for RT distributions of Car (left) and Face (right) choices in the V (top – in cyan) and AV (bottom – in pink) conditions. **c**, Regression coefficients (β) of the Early (light colours) and Late (dark colours) EEG component amplitudes (y 's) in A (cyan) and AV (pink) conditions, as predictors of the drift rate (δ) of the nHDDM shown in **a**. Dots indicate single-subject values and lines connect the population means. **d**, Across-subject correlation of differences in regression coefficients of the Late component (β_2 – x-axis) and differences in choice accuracy (y-axis) across conditions (AV-V). **e**, Boundary separation values (α) estimated by the nHDDM in A (cyan) and AV (pink) conditions. Dots indicate single-subjects and lines connect the population means. **f**, Non-decision times (τ) estimated by the nHDDM in A (cyan) and AV (pink) conditions. Dots indicate single-subjects and lines connect the population means.

Importantly, we deployed of a neurally-informed HDDM (nHDDM), whereby we incorporated single-trial EEG component amplitudes (y -values) into the parameter estimation (Figure 2.5a). Specifically, we extracted single-trial discriminator amplitudes from subject-specific temporal windows corresponding to both the Early (stimulus-locked) and the Late (response-locked) EEG components (see Materials and Methods). Since these values represent the amount of face or car evidence available for the decision (i.e. indexing the

quality of the visual evidence as we demonstrated in previous work (Ratcliff, Philiastrides and Sajda, 2009)) we used them to construct regressors for the drift rate parameter in the model, based on the idea that evidence accumulation is faster when the neural evidence for one of the choices is higher. We therefore estimated regression coefficients (β_{Early} , β_{Late}) to further assess the relationship between trial-to-trial variations in EEG component amplitudes and drift rate.

Our results revealed that the behavioural data were fit well ($R^2 = 0.94$) by the neurally-informed HDDM for both V and AV trials (Figure 2.5b). Consistent with the functional role of the two EEG components in conveying sensory and post-sensory evidence respectively, the within-subject single-trial discriminator amplitudes of both components were predictive of drift rate in both sensory conditions (Figure 2.5c; β_{Early} , β_{Late} significantly larger than zero for both V and AV, $t(39) = 19.25$, $t(39) = 15.56$ for $\beta_{\text{Early}}(\text{V})$, $\beta_{\text{Early}}(\text{AV})$ respectively and $t(39) = 15.36$, $t(39) = 20.66$ for $\beta_{\text{Late}}(\text{V})$, $\beta_{\text{Late}}(\text{AV})$ respectively, all P 's < 0.001). Furthermore, regression coefficients for the Late component were significantly higher than for the Early component in both conditions (Figure 2.5c, paired t -tests, $t(39) = -2.08$, $P < 0.05$ for V and $t(39) = -3.86$, $P < 0.001$ for AV) suggesting a higher modulation of the rate of evidence accumulation by the Late component amplitudes, consistent with the higher stimulus discrimination accuracy of this component and its role in encoding decision evidence.

Crucially, the contribution of the Late but not the Early component (i.e. β_{Late} , but not β_{Early}) was significantly higher in AV compared to V trials (Figure 2.5c, paired t -test, $t(39) = -3.30$, $P < 0.005$). This is consistent with the increased discrimination power of the Late component in AV trials and suggests that this component underpins the behavioural facilitation of evidence accumulation via post-sensory amplification of visual evidence entering the decision process. Furthermore, when we compared the differences in the component contributions across the two sensory conditions, we found that the difference between the Late and Early component amplitudes was significantly higher in AV compared to V trials (paired t -test, $t(39) = -2.09$, $P < 0.05$). This interaction effect further corroborates our conclusion that the addition of auditory information enhances the rate of evidence accumulation via post-sensory, and not early sensory, visual representations.

Next, we asked whether the observed drift rate increases in AV trials as explained by the Late EEG component (β_{Late}) were predictive of the behavioural improvements in accuracy across all 40 subjects. Indeed, we found a significant positive correlation (Pearson's $R = 0.41$, $P < 0.01$), suggesting that, overall, participants with greater amplification of their Late EEG component in AV relative to V trials, achieved stronger improvements in accuracy across the two conditions (Figure 2.5d). This result further validates the functional role of the Late EEG component in the observed behavioural benefits in AV trials via amplification of the post-sensory evidence entering the decision process itself.

We subsequently investigated the effect of the additional auditory information on the two other core parameters of the nHDDM. We found no difference in boundary separation between the two sensory conditions (Figure 2.5e) and significantly longer non-decision times in AV trials (Figure 2.5f, 370 ± 9 ms for V versus 408 ± 10 ms for AV, paired t-test, $t(39) = -4.68$, $P < 0.0001$). The latter indicates longer stimulus encoding in AV trials, since motor response production (indicated by the same button presses) should not differ between the two sensory conditions. This finding attributes the somewhat longer RTs we observed in AV trials (636 ± 16 ms for V versus 673 ± 18 ms for AV) primarily to longer stimulus encoding processes, which may result from the extra time required to process the auditory stimulus (see Discussion). Notably, the average difference in RTs (37 ms) is very similar as the average non-decision difference between the two conditions (38 ms), which provides further evidence for the early sensory origins of the longer RTs in AV trials.

2.4.4 Neurally-informed model outperforming behaviourally constrained model

Given that most previous studies in multisensory decision making have fit the drift diffusion models only to behavioural data, it is worth asking whether the inclusion of EEG-derived regressors actually improves model performance and/or shapes the conclusion derived from the model. We formally compared the neurally-inspired HDDM to a standard HDDM without neurally-informed constraints; the traditional model yielded a poorer trade-off between goodness-

of-fit and complexity (as assessed by the Deviance Information Criterion - DIC for model selection (Spiegelhalter *et al.*, 2002)) compared to its neurally-informed counterpart ($DIC_{HDDM}=2371$ vs $DIC_{nHDDM}=1865$). In addition, the conclusions that would have been derived from such a poorer model contradict those reported above. For example, the conventional HDDM yielded larger boundary separations in the AV trials (paired t-test, $t(39) = -3.52$, $P < 0.005$), the non-decision-times estimated by this model were ~100-120 ms longer for both sensory conditions compared to the nHDDM (490 ± 10 ms for V and 509 ± 11 ms for AV), and the difference in average non-decision times across conditions (19 ms) did not track the mean RT difference as closely as the non-decision times estimated by the nHDDM. Hence, this poorer performing model constrained only on the behavioural data could lead to the wrong conclusion that the auditory information also affects the response caution (or the speed-accuracy trade-off implemented by the subjects). This supports the importance of constraining behavioural models with neural data and suggests that integrating neural information in these models can potentially enable a more accurate characterisation of the behavioural effects as well as a mechanistic interpretation of their neural correlates.

2.5 Discussion

In this work we used multivariate single-trial EEG analysis and behavioural modelling to investigate the enhancement of visual perceptual decisions by complementary auditory information. We showed that significant improvements in behavioural performance in AV trials were accompanied primarily by enhancements in a Late EEG component indexing decision-related processes (Philiastides and Sajda, 2006a, 2006b, 2007; Ratcliff, Philiastides and Sajda, 2009). In contrast, an earlier EEG component encoding sensory (visual) evidence remained unaffected by the addition of complementary auditory evidence. Using neurally-informed cognitive modelling we showed that these multisensory behavioural and neural benefits could be explained primarily by improvements in the rate of evidence accumulation in the decision process itself.

2.5.1 Early and late accounts of multisensory decision making

There are two prominent theories in the field of multisensory decision making that emphasise either the role of early or late integration of multisensory information, respectively (Bizley, Jones and Town, 2016). The early integration hypothesis (Schroeder and Foxe, 2005; Ghazanfar and Schroeder, 2006; Kayser and Logothetis, 2007) posits that sensory evidence is combined at the stage of early sensory encoding. This hypothesis is supported by evidence for direct pathways between early visual and auditory regions or cross-modal influences on neural responses in early visual cortices (Ghazanfar and Schroeder, 2006; Eckert *et al.*, 2008; Wang *et al.*, 2008; Falchier *et al.*, 2010; Klinge *et al.*, 2010; Petro, Paton and Muckli, 2017) and studies demonstrating benefits for the perception of simplistic visual stimuli such as contrast (Talsma and Woldorff, 2005; Lippert, Logothetis and Kayser, 2007), motion direction (Esposito, Mulert and Goebel, 2009; Kayser, Philiastides and Kayser, 2017) and simple shape discrimination (Giard and Peronnet, 1999) from acoustic information. However, the use of such simple stimuli may have specifically engaged only early sensory regions, hence providing a biased interpretation that does not generalise to more complex objects.

In contrast, the late integration hypothesis proposes that evidence from each sensory modality is processed separately during early sensory encoding, and is combined into a single source of evidence downstream, during the process of decision formation itself (Bizley, Jones and Town, 2016). Support for this hypothesis comes from both animal and human experiments demonstrating that multisensory information is accumulated right up to the point of a decision, while processing of unisensory information occurs prior to the formation of a multisensory decision (Raposo *et al.*, 2012; Sheppard, Raposo and Churchland, 2013). Similarly, recent neuroimaging work has provided new insights that flexible behaviour can be accounted for by causal inference models (Körding *et al.*, 2007), with multisensory representations converging on higher-level parietal and prefrontal regions (e.g. inferior parietal sulcus, superior frontal gyrus) previously linked to the process of evidence accumulation (Heekeren *et al.*, 2004; Aller and Noppeney, 2019; Cao *et al.*, 2019; Rohe, Ehrlis and Noppeney, 2019).

Our findings appear to be at odds with the early integration hypothesis since we found no evidence that the addition of auditory information had any impact on the encoding of early visual evidence, which remained comparable between visual and audiovisual trials. Instead, we offered support of post-sensory enhancements of visual evidence with the addition of auditory information that is most consistent with the late integration hypothesis. Importantly, these later visual representations are likely to reside in higher-order visual areas involved in object recognition and categorisation (e.g. lateral occipital cortex), as we have shown previously (Philiastides and Sajda, 2007), consistent with a higher-level conceptualisation of the evidence (Aller and Noppeney, 2019). Specifically, the timing of these representations (starting after early sensory encoding and lasting until the commitment to choice) suggests that they unfold concurrently with the decision and provide the input to the process of evidence accumulation in prefrontal and parietal cortex (Heekeren *et al.*, 2004; Ploran *et al.*, 2007; Philiastides *et al.*, 2011; Filimon *et al.*, 2013).

2.5.2 Using neurally-inspired models to understand decision making

Crucially in this work, we were able to characterise the neural underpinnings of the behavioural benefits obtained from the addition of auditory information. This novel contribution was made possible by the joint cognitive modelling of behavioural and neural data that linked the neural correlates of sensory and decision evidence with the internal processes involved in decision making. Our neurally-informed drift diffusion model indicated that the improvement in behavioural performance derived mainly from an enhanced representation of post-sensory evidence that modulates the rate of evidence accumulation. This result ran contrary to the behavioural-only version of a standard drift diffusion model which provided a less parsimonious fit to the behavioural data and attributed the longer response times in audiovisual trials to additional changes (increases) in decision boundary and to a lesser extent in early encoding of the auditory stimulus.

We suggest that the reason for this discrepancy is a less accurate account of the trial-by-trial variability in the decision dynamics (also indicated by the poorer fit of the single-trial data) than its neurally-informed counterpart. In other words, the inclusion of the two well-characterised EEG components provided a more accurate account of the contributions of early sensory and decision evidence to the decision formation dynamics and thus enabled the disambiguation of the internal processing stage that yielded such a behavioural benefit. Additional support for this claim is provided by the fact that the behavioural model yielded longer stimulus encoding times whose difference across conditions did not track the difference in measured response times.

Our findings suggest that by constraining models of perceptual decision making, they can provide key mechanistic insights that may remain unobserved using behavioural modelling. This argument is in line with recent research suggesting that the high complexity of decision making models may yield neurally-incompatible outcomes (McGovern, Hayes, Kelly, & O'Connell, 2018; Turner, Gao, Koenig, Palfy, & McClelland, 2017; Turner *et al.*, 2015). However, when informed by neural measurements, these models can not only yield more reliable parameter estimates but also shed light on the neural mechanisms underpinning

behavioural effects (Ratcliff, Philiastides and Sajda, 2009; Cavanagh *et al.*, 2011, 2014; Ratcliff and Frank, 2012; Dmochowski and Norcia, 2015; Frank *et al.*, 2015; Nunez, Vandekerckhove and Srinivasan, 2017; Delis *et al.*, 2018).

It is worth noting that several previous studies have used DDMs to study multisensory decision making. Some of these considered models in which the combination of multisensory information was explicitly hard-wired, for example to converge during sensory accumulation (Drugowitsch *et al.*, 2012, 2014; Colonius and Diederich, 2018). By doing so, these models can describe certain aspects of human behaviour, but they can't evaluate competing hypotheses about the locus of convergence. Other multisensory studies have combined behavioural modelling using DDMs and EEG, but did not use the neural data to constrain the behavioural model. Using such an approach, we have previously argued that the encoding of visual random dot motion in early sensory regions is affected by acoustic motion (Kayser *et al.*, 2017), speaking in favour of a sensory-level integration effect. However, this sensory level effect was not validated using an EEG-inspired DDM model, as done here. One explanation for these diverging findings is that the use of simpler stimuli, such as random dot motion, may have biased the earlier study to a sensory-level effect, whereas multisensory information about more complex objects is instead combined at a post-sensory stage. This interpretation is supported by neuroimaging studies that have reported audiovisual interactions for complex stimuli mostly at longer post-stimulus latencies or in high-level brain regions (Beauchamp *et al.*, 2004; Stekelenburg and Vroomen, 2007; Werner and Noppeney, 2010a, 2010b).

Another, potentially important difference that might explain these divergent findings is the particular construction of the multisensory context across tasks. Many audiovisual integration studies use tasks in which there is a direct mapping between the source of the evidence across the two modalities, for instance, seeing a person's mouth while producing speech (i.e. lip reading), to compensate from sound loss in a noisy bar. In our task, as in many real-world scenarios, however, this direct audiovisual mapping is not immediately available. In our earlier example, the decision to cross the street on a foggy morning will be based on hazy objects in your visual field together with street sounds that cannot immediately be matched to individual objects. In other

words, the decision to step off the curb will be based on a broader audiovisual context and a higher-level conceptualisation of the evidence, such as the presence of car-like objects and sounds signalling a busy street. This is a subtle but critical distinction in deciphering the mechanisms underlying audiovisual integration and reconciling discrepancies across different experimental designs.

Though the Early component described in this study did resemble the well-described N170 component of the ERP in terms of the spatial distribution of activity, the temporal features were not exactly in alignment with it; by definition, the face-specific N170 occurs approximately 170 ms after stimulus presentation, however here we characterised our Early component as appearing around 220/230 ms post-stimulus. This could have been caused by a delay in early sensory encoding. The stimuli were presented very briefly compared to previous iterations of this task, making it more difficult for participants to process sensory information with less evidence present. It may have been harder to reach the stage of face-specific processing if there was less information to use to identify a face to begin with. Differences in non-decision time were also identified as linked to response time differences by our neurally-informed model.

On the contrary, the brevity of the sensory stimuli used here may have biased the results towards finding more significant effects in the later stages of decision making, as the short presentation provides very little to accumulate, which may place higher demand on the later decision stage. Further study might investigate this possibility by making small changes to the presentation time of the stimuli used in this study, and observe whether any early sensory processing effects begin to appear when evidence is presented for a longer period of time.

Finally, a limitation of this study is that it used congruent audiovisual stimuli throughout, meaning that auditory and visual stimulus content matched in what category of object/concept was present. This means that we are demonstrating the effect within intermodal integration, but not content-specific integration across modality. We cannot be sure that some of the effect we describe cannot be explained simply by the presence of additional sensory input. Interestingly, some studies have found improvements in visual motion perception even when additional auditory evidence, provided alongside visual stimuli, is not ‘useful’ to

the task, i.e. it does not provide any more information as to the direction of motion (Gleiss and Kayser, 2014). Another future direction of study using the current task may therefore be to investigate whether the same effect would have been observed if the additional auditory evidence was not useful to the task, meaning that sounds may be presented but they are not congruent with the visual stimulus category or do not help inform decision making. We might expect that unisensory (visual only) decision making performance would be lowest, followed by incongruent audiovisual trials, then congruent audiovisual trials. In conclusion, our analysis revealed significant enhancements to post-sensory decision processing that were associated with increased evidence accumulation. We successfully modified an existing discrimination paradigm as a framework to assess changes during audiovisual decision making, and employed a neurally-informed model to explain changes in behaviour associated with these enhancements. Some questions remained, such as whether a task with more direct mapping of information across modalities may provide a better opportunity to study evidence accumulation.

3 Chapter 3. Centroparietal positivity reflects audiovisual evidence accumulation

3.1 Summary

In our previous chapter, we revealed the temporal characteristics of enhancements to post-sensory decision processing during audiovisual decision making. The spatial representation of this effect resembled a centroparietal cluster. In addition, our results appeared to be at odds with the early integration hypothesis, revealing no significant changes during this period. However, the task used may have masked some of this effect, and further study of the temporal evolution of evidence accumulation was needed. Here, we asked participants to discriminate motion direction during a modified random dot motion task, adapted from Kelly and O'Connell (2013) to include visual and/or auditory information. An ERP analysis of electrophysiological activity within a centroparietal cluster (CPP) during the task revealed a clear pattern of gradually ramping activity leading up to the time of response. The rate of increase in this activity (i.e. evidence accumulation rate) was greater following presentation of audiovisual motion, and activity peaked at a higher level. Further, improvements in behavioural performance were partially explained by changes in the rate of evidence accumulation. This supported the findings of our previous chapter. However, our analysis revealed significant issues with the task design, which may have been capturing other non-decision processing such as detection activity, that requires further investigation.

3.2 Introduction

While many studies have investigated sensory processing and evidence encoding in a ‘clean’ environment, it remains important to observe how this process may change when a decision is required in the presence of external noise. Humans are frequently required to selectively collect sensory evidence from their surroundings in order to correctly respond to external cues. This often occurs in situations where sensory information from other sources makes this decision more difficult, but where it is crucial that the correct decision be made; someone about to cross the road on a foggy day needs to quickly and accurately judge whether they can see or hear any cars approaching, despite the poor conditions hindering their ability to do so. Being unable to properly sample evidence from this environment, or to successfully integrate complementary multisensory evidence, can impede your ability to make timely, accurate decisions. On the contrary, it is important that the process of sequentially sampling evidence from your environment facilitates decision making despite suboptimal conditions.

Some of the previous literature has therefore focused on the evidence accumulation process in experiments that simulate the presence of external sensory noise. In doing so we can observe, in a more realistic context, how relevant evidence from different sensory modalities is sampled and integrated to reach a decision boundary. In investigating the timing of sensory integration, a debate has formed between the early and late integration hypotheses (see Bizley *et al.*, 2016 for review). Some have argued that the integration of visual and auditory information takes place early on in a decision, and that it is tied to the initial encoding of sensory information. Indeed, several examples exist either describing early interactions between primary sensory cortices (Lakatos *et al.*, 2007; Chandrasekaran, Lemus and Ghazanfar, 2013) or relationships between early integration signals and behaviour (Iurilli *et al.*, 2012; Perrodin *et al.*, 2015). In opposition, others argue that sensory information is encoded largely in a unisensory fashion, and that any complementary information will not be combined until the later stages of decision formation (Werner and Noppeney, 2010b; Raposo *et al.*, 2012; Sheppard, Raposo and Churchland, 2013; Kayser, Philiastides and Kayser, 2017).

In Chapter 2, we began to address this by employing a ‘noisy’ audiovisual task with face/car stimuli designed to be more reflective of those in a real-world decision. We investigated how audiovisual integration might impact on decision evidence and found that, compared to a visual-only decision, complementary auditory evidence significantly enhanced the later post-sensory decision component and not the early sensory encoding component. This result was consistent with the late integration hypothesis. Neurally-informed modelling also found that this was reflected in the evidence accumulation component of the decision process, as compared to early sensory encoding. If audiovisual integration truly impacted on the later stage of the decision, we should see similar changes in the speed with which a decision boundary is reached.

Whereas this work revealed significant effects on the post-sensory evidence entering the decision, our next focus was therefore to investigate how this effect impacts on the decision process, and evidence accumulation, itself. In order to investigate the effect of audiovisual evidence on evidence accumulation, it is necessary to use a task that can properly capture this process. It would also need to isolate the evidence accumulation process itself from other activity related to presenting visual and auditory stimuli, such as visually evoked potentials caused by the sudden presentation of visual stimuli. These would potentially mask ramping activity and make any examination of changes difficult. Kelly and O’Connell (2013) describe employing a continuous monitoring random dot motion task, based off the prototypical direction discrimination task (Newsome, Britten and Movshon, 1989; Britten *et al.*, 1992). In this task, participants are presented with a central patch of moving dots, a set percentage of which are moving coherently in one of two directions. They are required to respond with the direction they believed the dots to be moving. In the 2013 paper, the task has been altered so that the intertrial intervals are replaced with the same style of moving dots, but with no coherent motion, rather than instantly transitioning from a blank screen to the dot stimuli. Participants are asked to continuously monitor the dots and respond when they perceive coherent motion, at the same time indicating the direction of motion they perceived. No cues were provided to indicate the start or end of the target motion period. Difficulty was modulated by changing the proportion of

coherently moving dots during the target motion periods, for a total of four difficulty levels.

Analysis of the associated electrophysiological response recorded as participants discriminated motion direction revealed a gradual ramping of activity, beginning around 200ms post stimulus and steadily increasing to peak just before the response. This analysis focused on a cluster of centroparietal electrodes that showed the maximum component amplitude, consistent with previous work (O'Connell, Dockree and Kelly, 2012), and this centroparietal positivity (CPP) signal was associated with the gradual build-to-threshold ramping activity likely to represent sequential sampling and evidence accumulation. The researchers found that the rate at which activity ramped up approaching the decision scaled with increased motion coherence, in other words with additional sensory evidence. Furthermore, the slope of the CPP predicted response times within coherence levels, suggesting a tangible link between neural enhancements and behavioural benefits. Interestingly, they also found that decision amplitudes peaked at similar levels just before the time of response, a peak they suggested might represent a common decision boundary-crossing stopping criterion.

We decided to use this work as a basis of investigation into the effects of audiovisual integration on evidence accumulation. As the existing task was developed with only a visual stimulus, we modified the task to include a complementary moving sound stimulus. Consistent with other works investigating audiovisual integration (Gleiss and Kayser, 2014; Kayser, Philiastides and Kayser, 2017), we employed a moving white noise stimulus. During random motion intervals, a simple, unmoving, white noise was audible over headphones. However, as the dots seamlessly transitioned from random motion to a target motion period, the sound began to move from a perceived central location to either the left or right, consistent with the direction of coherent dot motion, and gradually moved across during the target motion period. As in the original task design, we instructed participants to respond when they perceived coherent motion in either direction but did not cue our participants as to the start of a target period.

Our task was designed to present coherent, simultaneous audiovisual motion in this way (hereafter 'AV'). However, we also presented trials with dot (visual)

motion but not sound motion (hereafter ‘V’ trials), with the white noise remaining central (i.e. not moving), as our visual-only trials. To complement this, we also used sound-only trials where the sound would move during target periods, but dots would move incoherently throughout (hereafter ‘A’ trials). This allowed us to examine, should there be behavioural benefits or enhancement of either evidence accumulation or peak amplitude, whether those benefits were primarily driven by the visual or auditory information. This was something we did not account for completely in the previous chapter and sought to rectify in this task; AV and V trials were present but not A trials, preventing us from being able to make that same comparison when looking at temporal components.

Operating under the assumption that this task design would properly capture the ramping up of decision-related activity, we hypothesised that participants would discriminate motion direction more accurately on AV trials compared to V or A trials. In line with Kelly and O’Connell’s behavioural results, we also predicted that participants would miss fewer target motion periods in AV trials compared to both other conditions. Regarding the neural data, we first hypothesised that we would see an increased evidence accumulation rate during AV trials compared to V or A trials. With complementary information available, the quality of the decision evidence is enhanced above that of either unisensory trial type. This should be reflected in a steeper slope of the evidence accumulation rate in the ERP analysis.

When analysing the profile of neural activity around the response, there is some debate as to what to expect in the data specifically when locking it to the response time. Kelly and O’Connell found ramping activity reaching a common threshold around the time of response, something they attributed to a boundary-crossing stopping criterion, as discussed above. However, previous studies examined this time period found differences in peak amplitudes near the response, specifically modulation of said amplitudes with confidence. We calculated a confidence proxy as inversely proportional to the square root of the total decision time, and found that increased confidence positively correlated with deviations from the mean component amplitude at the time of choice on a trial-by-trial basis (Philiastides, Heekeren and Sajda, 2014). A further task allowed participants to reveal low decision confidence by opting out of a trial

for a small, certain reward. Analysis of EEG data revealed a gradual ramping activity likely reflecting evidence accumulation. On a trial-by-trial basis, fluctuations in the rate of accumulation predicted the likelihood of participants opting out (Gherman and Philiastides, 2015). Crucially, both these works did not find accumulation to a common amplitude threshold; both found that, dependent on trial-by-trial evidence strength or confidence, ramping activity did not meet a common threshold, instead reaching the time of decision at significantly different peak amplitudes.

Based on these findings and considering the differences in evidence quality between our conditions, we expected that differences in decision confidence could again lead to peak amplitude modulation. To investigate this, we conducted an exploratory analysis which investigated whether any amplification of the ERP at the conclusion of evidence accumulation was related to decision confidence, or further whether this amplification might be a measurable confidence index. Should the AV trials in fact be more confident decisions for participants, with enhanced decision evidence, these trials may have enhanced peak amplitudes compared to low confidence trials. We therefore hypothesise that AV trials will have higher peak amplitudes, around the time of response, compared to V or A trials.

3.3 Materials & Methods

3.3.1 Participants

Thirty-nine participants took part in the RDK EEG experiment. Five were subsequently removed from the analysis due to recording issues ($n=1$) and poor-quality data recording ($n=4$) during the EEG session. All results presented here are based on the remaining thirty-four participants (age range 19-33 years, 20 female). All were right-handed, and reported normal or corrected to normal vision, normal hearing, and no history of neurological problems. All participants were compensated at a rate of £6 per hour. This study was approved by the ethics committee of the College of Science and Engineering at the University of Glasgow (CSE 300150102) and informed consent was obtained from all participants.

3.3.2 Stimuli and task

The task design was based on an evidence accumulation study (Kelly and O'Connell, 2013), where a random dot kinematogram (RDK) stimulus smoothly transitioned between periods of coherent motion and random motion. Their task was designed with the intention of recording the isolated evidence accumulation processes during a perceptual decision. In other words, they intended to capture a less disrupted view of the neural ramping activity associated with evidence accumulation processes, importantly, without impedance from visually evoked potentials created through the sudden presentation of a new visual stimulus in each trial. In their task, large visual changes were kept to a minimum, while the motion information contained within the RDK stimulus changed. Participants were required to discriminate motion direction from moving dots in RDK stimuli that transitioned between coherent and incoherent (random) motion.

Using this task design, we intended to take a step further from unisensory perceptual decision making, to investigate audiovisual decision making, by introducing a new sound motion aspect. We presented this in the form of white noise that would move gradually from the centre either to the left or the right, simultaneously with RDK motion direction, and we asked participants to

discriminate motion direction having perceived this combined audiovisual evidence. We also presented either unisensory version, with RDK and white noise stimuli present in each trial, but with ‘useful’ coherent motion only present within one sensory modality depending on the trial.

Stimuli. All stimuli were displayed on an Asus ROG Swift PG278Q monitor (resolution, 2560x1440 pixels; 96x96 dots per inch; native refresh rate 144Hz, set to 120Hz). Participants were seated inside an electrostatically shielded room, positioned 75cm from the display, and instructed to remain still during each testing block. Experimenters monitored participants throughout the task for lack of attention, excessive movement, touching of the EEG equipment, or eye-closing, through a live webcam placed inside the testing booth. All participants were informed of this beforehand and gave their consent.

We created and presented all dot stimuli using the PsychoPy® software, version 1.82.02 (Peirce *et al.*, 2019). Stimuli consisted of random dot kinematograms (hereafter RDK; Newsome and Pare, 1988), whereby a proportion of the dots moved coherently in one direction (left vs. right), while the remaining dots moved in random directions. Specifically, each stimulus consisted of a dynamic field of white dots (number of dots = 118; dot diameter = 0.07 degrees of visual angle, dva; dot lifetime = 8 frames; dot speed = 0.96 dva/s), displayed centrally on a grey background through a circular aperture (diameter = 5 dva). We manipulated task difficulty by specifying the proportion of dots moving coherently in the same direction, with the remainder of dots moving in random directions (i.e. motion coherence). In the final design of the task, the three difficulty levels presented were of 22%, 33%, and 40% dot motion coherence. These were selected based on motion direction discrimination performance of near and above threshold during piloting of the task (see **Piloting** below).

We created sound stimuli using version 2.1.2 of Audacity® recording and editing software (Audacity Team, 2016). Static and moving sounds consisted of white noise generated within the program (amplitude = 0.5). Sounds moving from the centre to the left or right were created by modifying two independent stereo tracks using the pan and fade out functions within the program. One track, equally balanced between the left and right channels and so positioned centrally, would decrease in amplitude to zero at a consistent rate over the

duration of the sound. At the same time, we manipulated the balance between the left and right channels of a second stereo white noise clip, depending on the intended direction of motion and degree of motion of the final sound stimulus, and increased the amplitude of this over time as the first decreased. Coupled with the use of headphones, this created the perception of the sound moving from the centre to the left or right. We manipulated the degree of motion, or the distance by which the sound appeared to move from the centre over to the left or right (degrees), by adjusting the amount by which the second stereo track was weighted in either direction. For example, to create a sound that moved from the centre to 45° to the left, the balance of the second white noise clip would be set ('panned') 50% to the left, causing the sound to move at a steady rate during the target period from the centre to 45° to the left as the amplitudes of the first and second sound clips decreased and increased respectively. In the final design of the task, the three difficulty levels presented were of motion distances of 40.5° , 45° , and 67.5° . As with the visual stimuli, these were selected based on group average near and above threshold motion direction discrimination performance recorded during piloting of the task.

Piloting. To maintain task performance near threshold, here between 60% and 80% of responses correctly discriminating motion, we selected three visual and sound motion coherence levels based on the results of unisensory (one visual-only, one auditory-only) behavioural pilots of the task.

To pilot the visual portion of the task and to examine which levels of coherent motion were perceivable and therefore appropriate for use in the final experiment design, 15 participants were presented with periods of coherent and incoherent leftward or rightward motion using RDK stimuli. We manipulated dot motion coherence, i.e. difficulty, in 5% increments from 5-50% motion coherence. We presented a total of 320 trials in 8 blocks of 40 trials, with equal numbers of trials presenting each motion coherence level. Participants were asked to maintain focus on a central fixation cross, and monitor for the presence of coherent motion, responding with the direction in which they perceived said motion using the respective arrow keys (left and right only) on the keyboard. In a similar design and with the same number of trials presented, 8 participants completed a sound-only direction discrimination task with sound stimuli panning

from the centre to between 9-90° laterally, either to the left or the right. Participants were instructed to attend to the sound and, when they perceived the sound moving, indicate the direction of the perceived motion using the left and right arrow keys on the keyboard. They were asked to do so as soon as they perceived any coherent motion.

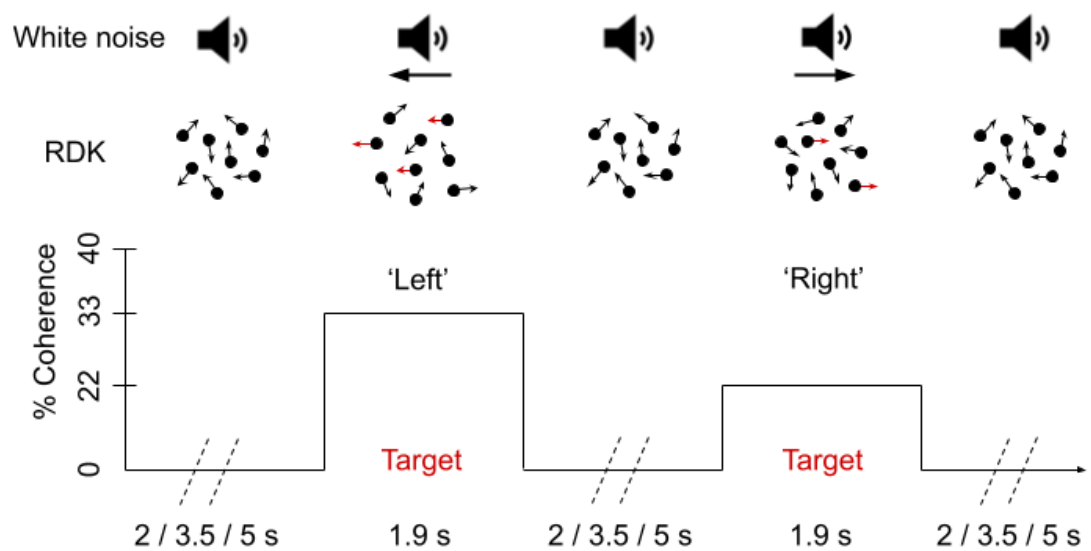


Figure 3.1 Behavioural task design. Figure adapted from Kelly & O'Connell, 2013. Participants were required to detect and discriminate coherent motion within audiovisual trials (as above) or unisensory (i.e. visual or auditory only) trials. White noise and RDK stimuli were presented throughout ITIs and target periods, however we presented coherent motion to the left or right during target periods only. No blank periods (i.e. without stimuli) were presented. RDK coherence was proportion of dots moving together in the same direction. White noise degree of motion was consistent during AV trials while RDK coherence varied (y-axis). On auditory trials, RDK would remain at 0% coherence throughout while white noise degree of motion varied between 40.5° / 45° / 67.5°.

Using the mean proportion of trials during which participants successfully discriminated motion direction, and extrapolating between the exact coherences presented, we selected three sound distances (40.5° , 45° , and 67.5°) and dot motion coherences (22%, 33%, and 40%) where motion discrimination was estimated to equate approximately 70%, 80%, and 90% accuracy. These were selected to present a range of difficulties from around threshold and above, or from relatively difficult to relatively easy.

Training. To familiarise participants with the white noise and dot stimuli prior, as well as to the lack of obvious cues for coherent motion periods during the main task, we asked them to complete two short training tasks. The first used a high level of motion coherence that we expected should be easily detectable, and we asked participants to respond when they perceived motion in either direction, until they responded correctly for five consecutive trials. This was usually completed within 10-15 trials. We then followed with another training task that included both visual and sound motion, this time with two levels of difficulty, one at a high motion coherence/distance that should have been relatively easy for participants to detect (but slightly more difficult than in the first training task), and a lower coherence/distance that we expected participants would miss some of the time (both based on early piloting). We also provided feedback that indicated if the participant had successfully detected a target motion period, or if they had missed one, using on screen text; participants saw “HIT” in green text if they responded during the target motion period, “MISS” in red if they responded during the ITI, and “TOO SLOW” in blue if a target motion period passed with no response. The intention of this training task was to familiarise participants with needing to continuously monitor for motion periods that may not be immediately obvious, as in the full task. This training task lasted ~3 minutes.

Due to the lack of obvious cues in the task, we found it could be relatively quick to tire participants, so we adopted a procedure of offering short, frequent breaks throughout the day of testing to minimise the effect of this on task performance or concentration. We also fitted the EEG cap on the participant

prior to training as allowing the gel time to settle tended to reduce the measured impedance levels by the time we began recording.

Main task. The main task was divided into 15 blocks of 36 trials each, for a total of 540 trials. Each block exclusively presented visual, auditory, or audiovisual motion, and lasted 3 minutes 15 seconds on average. There were 5 blocks per modality, with the order of block presentation pseudorandomised at the beginning of each participant's task using the PsychoPy shuffle function. An equal number of the 540 trials presented either A, V, or AV motion, each of the three difficulty levels, and leftward or rightward motion, for a maximum of 30 repetitions of each unique trial type (60 not including motion direction). This was to ensure sufficient power during statistical analysis of the results. While sensory modality was manipulated between blocks, task difficulty and motion direction changed within each block, with an equal number of trials of each difficulty level and direction presented in each block.

We provided 30 second breaks between each block to prevent fatigue, and after the fifth and tenth blocks participants were provided with an untimed break which they could end once they were ready to continue; this was usually only a few minutes long. Subjects were allowed to get up out of their seat if requested, but this was not suggested or encouraged as we did not want to unnecessarily disturb the EEG equipment. We cued participants with the modality of each block from five seconds prior using on-screen text; after a break, this also served as a warning as to the beginning of the next block.

Participants fixated on a central cross for the duration of each block and were instructed to maintain attention to both visual and sound stimuli regardless of block category. This was specified due to some participants employing tactics during piloting such as choosing to close their eyes during sound motion blocks. They were instructed to monitor the stimuli for motion towards either the left or right, and that they were to indicate as soon as they saw leftward or rightward motion using the relevant index finger and arrow key on the keyboard (left index finger on left arrow for leftward motion, and vice versa for rightward motion). No feedback was given during the task, and coherent motion periods would continue to present for their full length regardless of when a keypress was detected, so as to prevent a) unnecessary cues signalling the change between

coherent and incoherent motion, and b) additional VEPs potentially caused by the small change in audiovisual information influencing neural data recorded around the time of the response.

During V motion blocks, dot motion coherence alternated between complete incoherence (0 during intertrial intervals (ITI), and coherent motion (one of either 22%, 33%, or 40% coherence) during the target motion period, however no auditory motion was present, with a static white noise of equal amplitude presented throughout. In the case of A motion blocks, the opposite was true; the auditory stimulus (white noise) alternated between being static during the ITI, and transition to one of three motion distances (40.5° , 45° , and 67.5°) during the target motion period. Finally, for AV motion blocks, stimuli alternated between static/incoherent motion during the ITI and coherent motion during the target period; in this case specifically, we presented all three dot motion coherences but maintained sound motion at 45° only. At the end of each target period, the next ITI began, and so incoherent or static motion was presented as defined by the modality of the block. The transition between the ITI and target motion period in each case was seamless; without any clear cue or disruption of presentation, the dot stimuli switched from incoherent motion in one frame to coherent in the next, and regarding the white noise stimulus, from a static ‘central’ sound to one moving gradually from the centre to the left or right.

3.3.3 Behavioural analysis

To calculate differences in task performance, we used the timings and condition codes provided by EEG recording events. This ensured that all results translated accurately between behavioural and neural analyses. At an earlier stage of analysis during piloting, we attempted to take the accuracy of motion direction discrimination as a measure of task performance. Specifically, we calculated performance accuracy as the proportion of trials where participants correctly discriminated the direction of coherent motion presented in the trial. In doing so however, we found widespread ceiling effects (see **Behavioural results**). Instead, differences in task performance seemed to be embodied by their hit rate, that being the proportion of trials where participants responded during the

target motion period, or specifically trials where participants were confident enough that they could perceive coherent motion that they responded. For this reason, further analysis of task performance focused on hit rate (coherent motion detection) rather than direction accuracy (motion direction discrimination).

We subsequently calculated the mean miss rate, specifically the proportion of trials where participants failed to respond during the target motion period, per participant, and the group mean miss rate per condition (modality and difficulty). To evaluate variability, we calculated the standard error of the mean for each condition.

To compare participant response speeds across conditions, we calculated the median response time per participant, per condition, and the mean across the group per condition. We then calculated the standard error of the mean across the group.

3.3.4 EEG data acquisition

Participants performed the training tasks and main task on the same day, in a dark, sound-attenuated, and electrostatically shielded room. Continuous EEG data were recorded using a 64 channel EEG amplifier system (BrainAmps MR-Plus, Brain Products, Germany) with Ag/AgCl scalp electrodes placed according to the international 10-20 system on an EasyCap (Brain products GmbH). Channels were referenced to the left mastoid, with a chin electrode acting as ground. During recording set up, we ensured input impedance was $<30\text{k}\Omega$ in all cases, with most $<20\text{k}\Omega$, and sampled the data at a rate of 1000 Hz at an analogue band pass of 0.0016-250Hz. To obtain accurate event onset times, experimental event codes and participant responses were recorded simultaneously with the EEG data using PsychoPy and Brain Vision Recorder (Version 1.10, Brain Products, Germany).

As well as the task, participants completed an eye-movement calibration task; they were asked to blink naturally at a white fixation cross on a grey background

for a period of ten seconds, before following the cross in lateral and vertical saccades around the screen. The exact timings of blink and eye movement events were recorded in the same method as task events, to be used in EEG data pre-processing.

3.3.5 EEG data pre-processing

The data were pre-processed offline (excluding the previously mentioned analogue band pass, see EEG data acquisition) using MATLAB (The MathWorks, 2015b). We applied a 0.5-40Hz bandpass filter to remove slow DC drifts and higher frequencies, to focus our analysis on slower evoked responses. We also used data recorded in the eye-movement calibration task to identify linear artefacts associated with eye blinks and eye-movement, using a principal component analysis approach as described in (Parra *et al.*, 2005) to remove them. The data were re-referenced to the average of all channels and baseline corrected to 100ms prior to stimulus onset. Following eye-movement removal, further trials were removed if participants responded so soon after the onset of coherent motion that it was unlikely they were responding to this (i.e. trials where the response time was faster than 300ms post-stimulus), or where amplitudes exceeded 150 μ V.

3.3.6 EEG data analysis

We analysed stimulus- and response-locked ERPs to identify temporal activity related to accumulation of perceptual evidence, specifically of visual and/or auditory neural evidence of coherent motion. ERP analyses were performed using custom MATLAB code and the Current Source Density (CSD) toolbox (Kayser and Tenke, 2006). We sectioned the pre-processed data into stimulus-locked epochs, from -750ms to +2000ms relative to the onset of coherent motion of any condition. We then converted this data to current source density, using the CSD toolbox, which aims to minimize the influence of volume conduction while increasing spatial selectivity. This step was also performed to provide

consistency with Kelly & O'Connell (2013), who also applied this transformation. We applied this filter to the data of each participant individually, before proceeding with the analysis.

We then transformed our stimulus-locked data into shorter response-locked epochs (-600ms to +100ms) using individual trial response times. For each trial, we found the response time, took this time as 0ms, then created our new epochs around this. We calculated the mean (ERP) across participants and coherence levels, per condition, ignoring NaNs, followed by calculating the mean across participants for the group mean.

We created a separate script within MATLAB to analyse the amplitudes and ramp of activity within these ERPs. First, we specified a group of six posterior electrodes across which to focus our analysis (CP_z, P₁, P₂, PO₁, PO₂, O_z ; see Figure 3.3). Sensors with the steepest increase in activity approaching the time of response, collapsed across all conditions, were selected for our analysis. Interestingly, the spatial distribution of steeper slopes approaching the point of decision was similar to those of Kelly & O'Connell's work, where they identified the centroparietal positivity component linked to evidence accumulation and decision formation. That the ramping activity was similar to that identified by Kelly & O'Connell gave us confidence that we had successfully replicated and expanded on their decision making task, here adding the auditory element without compromising the overall effect of observing evidence accumulation. Therefore, we continued with this selection of sensors to explore the effect of additional auditory evidence on the visual-only decision making described by the previous authors, more specifically looking at its effect on evidence accumulation approaching the time of decision.

We then endeavoured to quantify, and to compare between conditions, the temporal peak and rate of evidence accumulation during coherent motion periods. To define the rate of evidence accumulation in stimulus-locked data, specifically the rate of increase in activity over time in our ERP, we visually selected the beginning of amplitude increases, based on ERPs averaged across the group, and the approximate timepoint of the peak amplitude across the group, for each condition, which formed the start and end points of our analysis windows. We therefore defined condition-specific time ranges that would

encompass this whole period (+300ms to +1100ms for sound-only trials, +200ms to +700ms for visual-only and audiovisual trials). We then found the maximum value within these time ranges, per participant and condition. For response-locked peaks, we started our range at -550ms using the same method and found the maximum ERP value between -150ms to +50ms around the response, anticipating that the majority of peaks would fall just before the response. We then fit a linear regression to each subject-specific and condition-specific ERP within these predefined time windows using custom code in MATLAB, using m , the slope, as our measure of rate of evidence accumulation. Finally, we compared the rates of evidence accumulation (m) and peak amplitudes of each condition using (as appropriate) parametric or non-parametric analyses of variance, and post-hoc pairwise comparisons

3.4 Results

3.4.1 Behavioural results

For this task, our intended method of measuring behavioural performance, or specifically to discriminate dot motion direction, was to take the proportion of correctly discriminated trials as a proportion of the total number of trials, i.e. accuracy. However, during our pilot experiments we found that this was unlikely to be a useful measure of performance. In separate visual and auditory pilot tasks implemented to verify coherences with perithreshold performance, we found that participants would frequently present very high decision accuracy once missed trials were excluded (98.22%, 97.68%, and 97.76% median accuracy in V, A, and AV trials, respectively). After increasing the motion coherence ranges in subsequent pilots for each condition, we found that, while accuracy did scale with increasing motion coherence, this seemed to be almost completely driven by whether participants responded at all; when excluding missed trials, participants almost always correctly discriminated motion direction.

Interestingly, Kelly & O'Connell themselves reported that miss rate, rather than direction discrimination accuracy, was modulated by sensory evidence; participants correctly discriminated motion direction in more than 98.5% of trials that they responded to in time. It appeared that we had captured the same effect due to the same ceiling effect just described, as did a more recent work using the same paradigm (Newman *et al.*, 2017; Van Kempen *et al.*, 2019). Following this, we continued to calculate miss rate as our measure of behavioural performance, both so that we would hopefully be able to see decision behaviour modulated by our conditions, but also to stay consistent with the original work.

This apparent inability to use accuracy as an effective behavioural measure, in either the previous paper or this experiment, may suggest that the behavioural task failed to properly capture decision processing. If participants were able to correctly discriminate motion in almost all trials where they successfully detected that a target motion period is occurring, or that coherent motion was present, this may suggest that this task captures more of a detection process

rather than true discrimination. If this task were capturing motion discrimination, we should be able to see gradual increases in accuracy with increased motion coherence, independently of the miss rate. However, it is possible that this task would not fully capture evidence accumulation ramping up to the decision as intended, as it seems more demanding of successfully detecting motion and not discriminating it. Van Kempen *et al.* (2019) argues that this effect was because the motion coherence levels used were above threshold, and a measure of performance variability, calculated by dividing the standard deviation in RT by the mean, was used instead. However, our lowest motion coherence was selected based on pilot data, where participants should have been able to discriminate motion direction in ~70% of trials, without excluding missed trials. This suggests that the ceiling effect we saw once removing missed trials may be related to whether participants were primarily having to perform detection or discrimination task, rather than the task being too easy to allow meaningful comparisons of discrimination accuracy modulation. For this reason, we approached the rest of our analysis with caution.

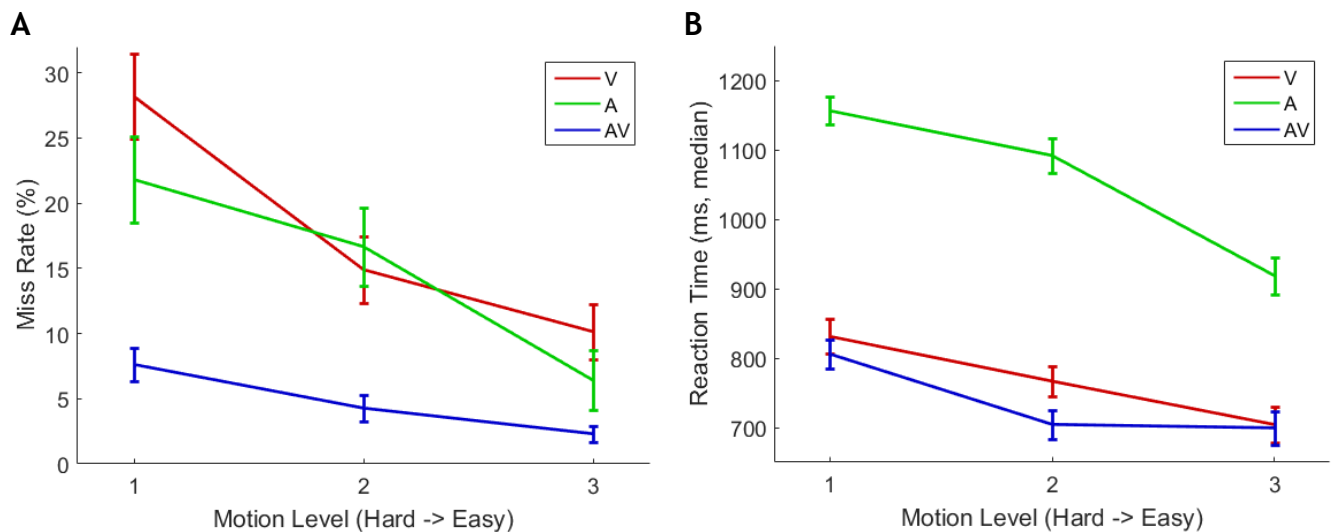


Figure 3.2 Behavioural results of audiovisual motion task. Colours represent different sensory conditions; red = visual trials (V), green = auditory trials (A), blue = audiovisual trials (AV). Motion level (hard -> easy) represents increasing motion coherence (dots) or distance (sounds; see **Stimuli and task**). Error bars are standard error of the mean. **A** Miss rate of responses, meaning the proportion of trials in which participants failed to respond by indicating any direction of motion during the target motion period. **B** Media response latency per condition and difficulty level, from the start of target motion period. Averaged across trials and participants.

Miss rate. First, we inspected the rate at which participants failed to respond during target motion periods (miss rate, Figure 3.2a). It is clear that, at the most difficult motion coherence, participants missed coherent AV motion (mean = 7.63%) much less often than A or V trials (21.81% and 28.19% respectively). This difference remained true as motion coherence increased, however the difference between AV trial miss rate and the unisensory conditions did reduce at higher motion coherences (2.31%, 6.42%, and 10.15% respectively at the highest level).

We conducted a repeated-measures ANOVA to assess the effects of modality and coherence on miss rate. Regarding modality, Mauchly's test indicated that the assumption of sphericity was violated ($\chi^2(2) = 15.65, p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .72$). The results showed that there was a significant effect of whether coherent motion was present in the dots and/or sounds on miss rate, and with a large effect size indicated by partial eta-squared estimates ($F(1.44, 47.59) = 12.21, p < .001; \eta_p^2 = .27$). For coherence, Mauchly's test indicated that the assumption of sphericity was violated ($\chi^2(2) = 40.49, p < .001$), and we again corrected degrees of freedom using Greenhouse-Geisser estimates of sphericity ($\epsilon = .58$). The results showed that there was a significant effect of the amount of evidence of coherent motion on miss rate, with a large effect size ($F(1.16, 38.42) = 66.94, p < .001; \eta_p^2 = .67$). Finally, investigating an interaction between modality and coherence, Mauchly's test indicated that the assumption of sphericity was violated ($\chi^2(9) = 27.94, p = .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .74$). This was also highly significant, with a strong interaction between coherence and modality evident ($F(2.98, 98.84) = 24.13, p < .001; \eta_p^2 = .42$).

In summary, while miss rates for A and V trials were relatively comparable, the amount of motion evidence presented, and whether motion was presented visually, aurally, or audiovisually, all had significant effects on whether participants were able to successfully respond during the target motion period. Specifically, when motion was presented audiovisually, participants were more likely to correctly detect coherent motion.

Reaction time. Next, we investigated the effects of modality and coherence on reaction time (Figure 3.2b). From plotting the resulting data, the most prominent effect was clearly that the auditory condition had considerably longer reaction times than either of the other two conditions. However, this may be explained by the extra time needed for participants to detect motion due to the design of the stimulus (see Discussion). Other than this difference, we found that V and AV trial reaction times were comparable, with all modalities trending towards faster reaction times with increased motion coherence.

We conducted a further repeated-measures ANOVA to assess the effects of modality and coherence on reaction time. Regarding modality, Mauchly's test indicated that the assumption of sphericity was violated ($\chi^2(2) = 16.86, p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .71$). The results showed that there was a significant effect of whether coherent motion was present in the dots and/or sounds on reaction time, with a large effect size ($F(1.42, 46.82) = 123.99, p < .001; \eta_p^2 = .79$). For coherence, Mauchly's test did indicate that the assumption of sphericity was violated ($\chi^2(2) = 7.12, p = .028$), and degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\epsilon = .87$). The results showed that there was a significant effect of the amount of evidence of coherent motion on reaction time, with a large effect size ($F(1.75, 57.60) = 228.57, p < .001; \eta_p^2 = .87$). Finally, investigating an interaction between modality and coherence, we again corrected degrees of freedom with Huynh-Feldt estimates of sphericity ($\chi^2(9) = 17.20, p = .046; \epsilon = .87$). This was also highly significant, with a strong interaction between coherence and modality evident ($F(3.47, 114.53) = 16.69, p < .001; \eta_p^2 = .36$).

In summary, both the amount of sensory evidence and the modalities in which they were presented had significant effects on reaction time, reducing them in both cases.

3.4.2 EEG results

To measure the rate of evidence accumulation during the course of the decision, we computed an event-related potential (ERP) driven analysis, calculating the slope of a linear fit to ramping activity leading up to the time of response as our measure of evidence accumulation rate (see **EEG data analysis**). A steeper slope, meaning an increased rate of change in activity over time, would indicate a faster rate of evidence accumulation. The spatial distribution of sensors where activity ramped up gradually near the time of response can be seen in Figure 3.3, which highlights the cluster of sensors selected for inclusion in the ERP analysis (using data collapsed across conditions).

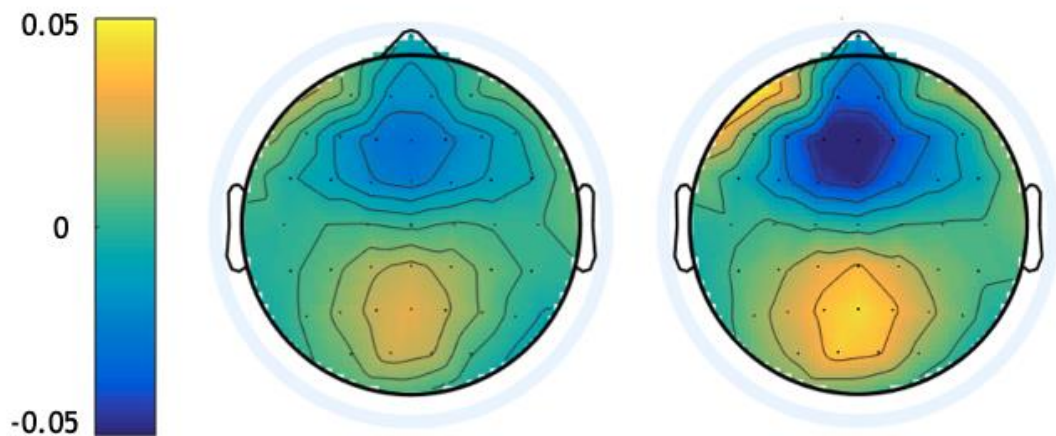


Figure 3.3 Topography of ramping activity, collapsed across conditions. Spatial distribution of rate of increase in activity during period approaching point of decision, using the slopes of sensor-specific linear regressions calculated across all conditions, between 100ms and 800ms post stimulus onset (**left**) or between -500ms and 0ms pre-response (**right**). Lighter/more yellow colouring represents regions where activity was more steeply increasing approaching the point of decision, darker/more blue regions represent activity more steeply declining approaching the point of decision. Regions where activity most steeply increased during the trial, namely the six centroparietal electrodes (CPz, P₁, P₂, PO₁, PO₂, Oz) circled by the inner-most ring on the figure, corresponded approximately with those identified as the centroparietal positivity component by Kelly and O’Connell (2013), providing reassurance that we had captured the same component in replicating and modifying their task. We selected the six sensors included in the rest of the EEG analysis based off of this information.

First, we investigated the effects of sensory modality on the rate of evidence accumulation when centring ERPs on the point of stimulus presentation. We collapsed across motion coherences to increase the number of trials per modality and increase statistical power, and plotted separate ERPs per condition (V, A, and AV trials). We then fit linear regressions to each condition between predefined start points and the peak amplitude of each condition, chosen from inspection of ERPs averaged across conditions (see **EEG data analysis**), in order to use the slope of each regression to quantify the rate of change in activity.

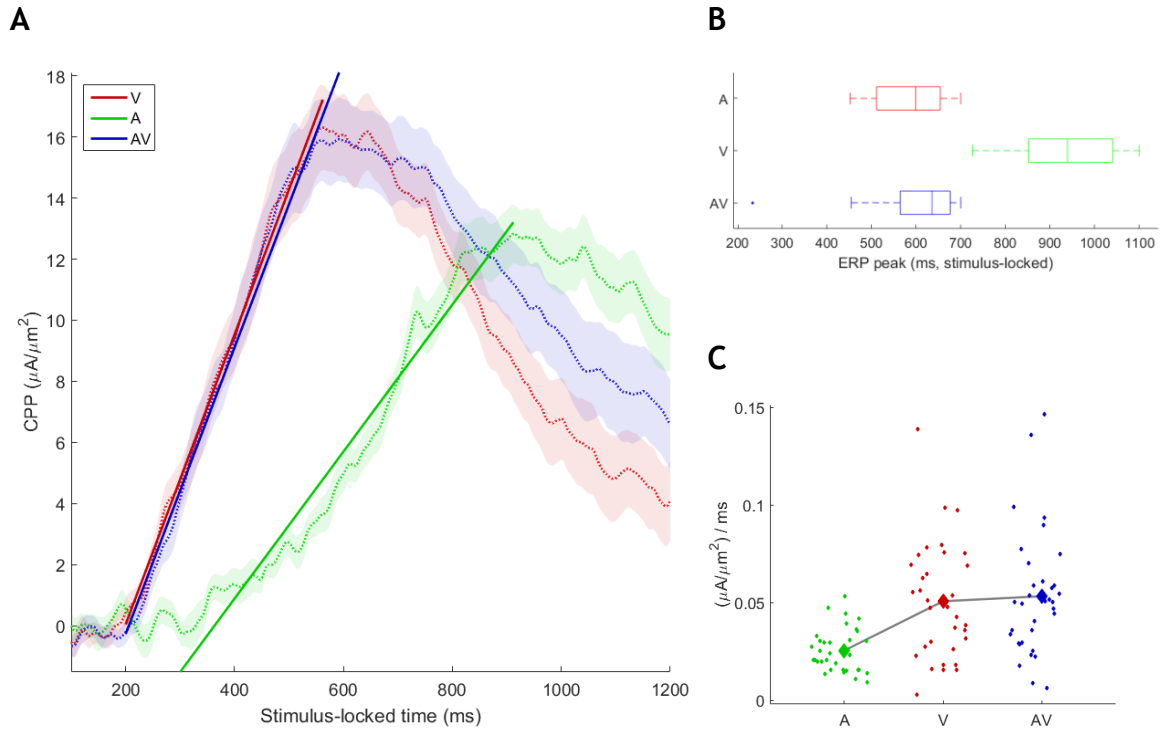


Figure 3.4 Temporal profile of stimulus-locked ERP from CPP sensors. Colours indicate conditions as follows: red = visual trials, green = auditory trials, blue = audiovisual trials. **A** Dotted traces represent group-averaged (mean) ERPs of activity measured from the CPP cluster from the onset of coherent motion presentation, i.e. the start of the target motion period. Trials averaged across participants and coherences, per condition. Shaded areas surrounding each trace represent standard error of the mean. Straight solid lines are linear regressions fitted to each condition, starting at 200ms (V & AV) or 300ms (A) post-stimulus, until the mean time of subject-specific peaks per condition. **B** Box plot representing the distribution of peak times selected for each participant per condition. Centre line of each boxplot is median peak time. Peaks were selected based on code searching for maximum response amplitudes between 200ms and 700ms (V & AV trials) or 300ms and 1100ms (A trials only). **C** Slope of linear regressions fit per participant and condition, with group median (large diamond) and individual slopes (scatter).

On initial inspection of the stimulus-locked ERPs, the ramping of activity of V and AV trials appeared considerably steeper than that of A trials (Figure 3.4). V and AV trial activity increased sharply from around 200ms post-stimulus, increasing at a consistent rate and peaking just before 600ms post-stimulus. When combined with the average response times seen in these two conditions, it appears that, should the activity from these electrodes be measuring evidence accumulation as hoped, that this peaked roughly 200-250ms before participants responded. The most significant period of decrease after this peak also appeared to begin as the average point of response was reached. However, as this analysis used stimulus-locked data, the actual distribution of response times represented here will spread considerably around the condition-averaged peaks seen in this figure, and so it is hard to pinpoint exact differences in timing from this visualisation alone. The only discernible difference between stimulus-locked AV and V trial ERPs was that the two slowly diverged as the trial progressed, with AV responses remaining slightly higher overall compared to V.

Interestingly, A trial activity did not begin to ramp up until slightly later, just after 300ms post-stimulus, and continued to build until much later in the trial, over 900ms after stimulus presentation. Activity seemed to increase at a slower rate, more gradually building up from the pre-stimulus baseline level to one approaching (but not matching) those of AV and V trial amplitudes (Figure 3.4c).

We qualitatively assessed the distribution of subject-specific, stimulus-locked slopes within each condition using the Kolmogorov-Smirnov test, and found that slopes fit to the average of AV trials were significantly non-normal ($D(34) = 0.17$, $p = .012$). There were also two notable outliers within AV trial data and one within V trial data, and a significant level of skewness in steepness of both V ($p < .05$) and AV trial ($p < .01$) slopes. As normally-distributed data are assumed to use a parametric approach, we used non-parametric methods to examine the differences in our data. Doing so allowed us to account for the non-normal distribution of the data and still make a fair assessment of any significant effects present, but at the cost of the additional statistical power garnered by parametric tests.

We statistically assessed the difference in the rate of evidence accumulation of each condition using Friedman's ANOVA, the non-parametric equivalent of a

repeated-measures ANOVA. Specifically, we compared the steepness of subject-specific linear slopes, fit to the mean ERP for V, A, or AV trials, to assess whether the sensory modality of motion was a significant factor affecting the steepness of the slope. Sensory modality was found to be a significant factor affecting the steepness of the slope ($\chi^2(2) = 27.941, p < .001$). Wilcoxon tests were used to follow up this finding, to further understand differences between individual modalities. A Bonferroni correction was applied; therefore all effects are reported at a .0167 level of significance. Wilcoxon test statistics are reported as z-scores (absolute values). A trial slopes ($Mdn = 0.0234$) were found to be significantly shallower than both V trial ($Mdn = 0.0488; z = -4.47, p < .001$) and AV trial ($Mdn = 0.0501; z = -4.49, p < .001$) slopes, with medium to large effect sizes ($r = -0.44$ in both cases). This suggested that the rate of evidence accumulation was significantly slower during trials with only auditory motion evidence compared to those with visual motion, either instead of auditory motion or in addition to it. There was no significant difference between the steepness of slope during V and AV trials ($z = -.214, p = .840, r = -0.02$), suggesting comparable rates of evidence accumulation where visual motion evidence was present.

This result mirrored our findings that A trial response times were longer than those in V or AV trials; it would make sense for evidence to accumulate over a longer period, therefore showing a shallower slope, if the decision was reached later in the trial. As A trial responses were considerably longer than those during AV or V trials, therefore occurring much later into the trial past the ‘lock’ point of this ERP, it is possible that a greater amount of variance in the time course of the response may have been ‘spreading’ the ramping of activity we had tried to visualise, making it more difficult to observe clearly any features of responses taking place later than others. In fact, when comparing the peak times of ERPs identified by our code (i.e. the times taken as end-points when we fit linear regressions to subject-specific data) between conditions using repeated-measures T-tests, we observed that the A trial ERPs peaked significantly later in the trial compared to V ($p < .001$) or AV ($p < .001$) trials. Figure 3.4b visualises this difference, as well as the relatively high variance in A trial peak times compared to V or AV trials.

In summary, A trial ERPs tended to peak further into the trial and became more variable in their timings, when starting from the time of stimulus presentation, than A or AV trial responses. This suggested that, in order to best compare the rate of evidence accumulation between conditions, locking to the time of stimulus presentation may not be most appropriate, as the effects of later responses weighed on the temporal profile of the data. In order to account for the greater variability and more carefully examine these slopes, we realigned the data around the time of response, specifically the point at which participants submitted their choice of motion direction via a button press. This meant that we could more directly compare the slopes of each condition, and more clearly capture the final few hundred milliseconds of activity as participants made their decision and acted on it.

Once we had aligned our data to the time of response and visualising it, it became immediately apparent that a much clearer comparison could be made (Figure 3.5). Whereas previously the A trial ERPs had appeared significantly different to V and AV trials, we believe largely due to the comparatively extensive response times, we could now see the temporal characteristics of activity in the period immediately preceding the response more clearly, and therefore the point of decision being made.

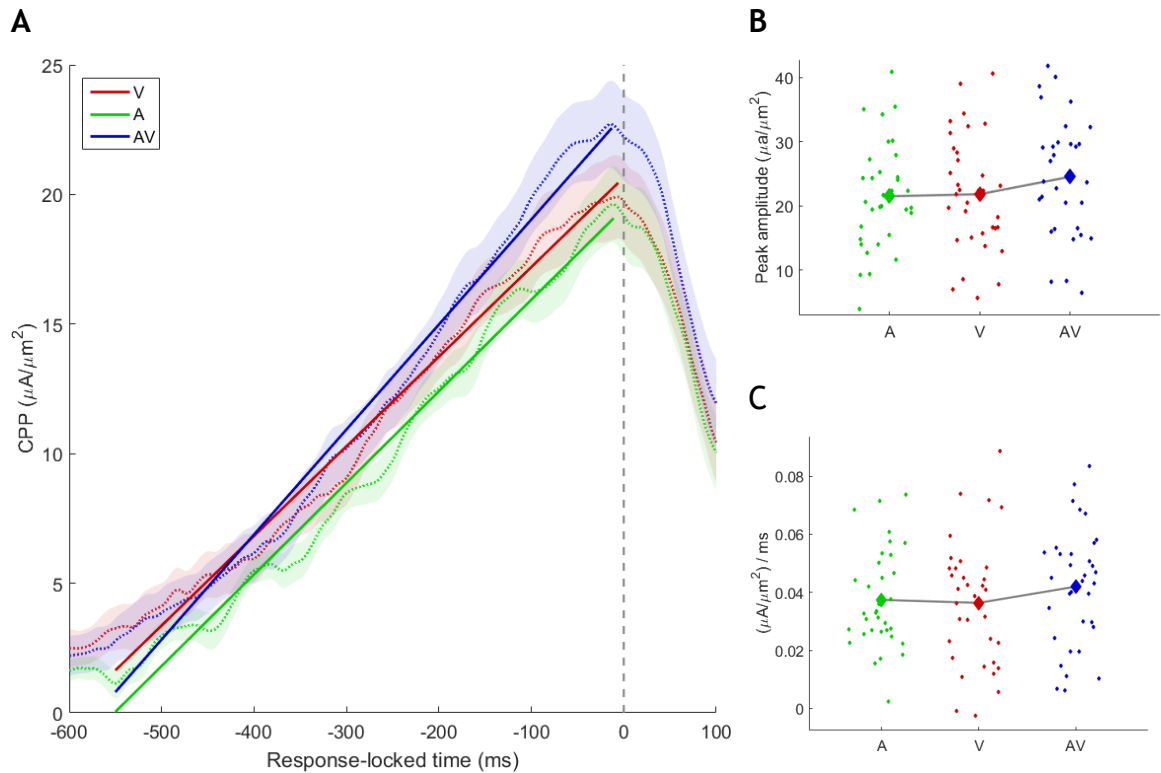


Figure 3.5 ERPs locked to the time of response, per sensory condition. Colours indicate conditions as follows: red = visual trials, green = auditory trials, blue = audiovisual trials. **A** Dotted traces represent group-averaged (mean) ERPs of activity measured from the CPP cluster, centred on the point of response in each trial (i.e. response-locked). Trials averaged across participants and coherences, per condition. Shaded areas surrounding each trace represent standard error of the mean. Straight solid lines are linear regressions fitted to each condition, starting at -550ms pre-response, until the mean of subject-specific peaks selected per condition between -150ms and +50ms around the response. **B** Median peak ERP amplitude near the point of response, across participants and coherences (large diamond), along with single-subject data (scatter). **C** Slope of linear regressions fit per participant and condition, with group median (large diamond) and individual slopes from subject-specific fits (scatter).

The change in amplitude between conditions was relatively similar, with all three slowly ramping up to a peak that appears to be just before the time of response. There appeared to be some differences in the rate of increase between conditions; while A and AV trial slopes started at similar points, by the time the point of decision was reached they had diverged considerably.

A Kolmogorov-Smirnov test found that the data were normally distributed (all conditions $p > .05$). There were no significant outliers within any of the conditions, and no significant levels of skewness or kurtosis. The data therefore met the initial assumption of being normally distributed to qualify for parametric statistical tests, and we chose to proceed with using a repeated-

measures ANOVA as a suitable assessment of any significant main effects. Mauchly's test for sphericity returned as significant and so the assumption of sphericity was violated ($\chi^2(2) = 28.81, p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .63$). The results found no significant main effect of sensory modality on the steepness of slope ($F(1.26, 41.42) = 1.88, p = .160$). This suggested that sensory modality had no effect on the rate of evidence accumulation between conditions.

As a further check, and as the assumption of sphericity had been violated, we next assessed the extent of the effect of sensory modality on slope steepness (i.e. evidence accumulation) using the less stringent but less statistically powerful Friedman's ANOVA, as we had when statistically analysing stimulus-locked ERP data. Although we had attempted to correct for the violation of sphericity using the Greenhouse-Geisser correction, it is possible that it was still clouding any effects in our data. A non-parametric test requires fewer assumptions to be met, but generally does not have the statistical power of its parametric counterparts and so should not return significant results where only the strength of the effect is of issue; it should only reveal a significant effect where other issues in the distribution of the data make parametric tests unsuitable for use. In this case, a Friedman's ANOVA found a significant main effect of sensory condition on the steepness of ERP slopes ($\chi^2(2) = 8.18, p = .016$). We followed up by assessing differences between individual conditions using Wilcoxon tests, with a Bonferroni correction applied, meaning all effects are reported at a .0167 level of significance. We found that AV trial slopes ($Mdn = 0.0445$) were significantly steeper than V trial slopes ($Mdn = 0.0399; z = -2.90, p = .003$) with a medium effect size ($r = -0.28$). Contrary to our stimulus-locked analysis where all significant differences were between A trial slopes and those of the other two conditions, these results suggest that, when centring our analysis on the point of decision, the greatest change in the steepness of ERP slopes was between trials with both auditory and visual motion evidence and those with visual evidence alone.

In summary, AV trial slopes were significantly steeper than in V trials, suggesting that the rate of evidence accumulation was increased during this condition, potentially due to the complementary auditory motion information present in

this condition. Interestingly, though they appeared to be different, there was no statistically significant difference between AV and A trial slopes (A trial $Mdn = 0.0329$; $z = -1.74$, $p = .084$, $r = -0.17$). By realigning the ERPs of the earlier stimulus-locked analysis to the time of response, we revealed that the rates of evidence accumulation were comparable when participants were presented with motion evidence of one sensory modality alone. In fact, the boost to behavioural performance seen during AV trials, in the form of decreased miss rates, seemed to coincide with the steeper slope we observed. This suggests that evidence accumulation occurred at an increased rate during trials with audiovisual motion, with the difference being the amount of sensory evidence for coherent motion present, and indeed the presence of complementary audiovisual evidence of coherent motion.

Another interesting difference visible in the profile of ERP activity preceding the time of response (Figure 3.5) was in the peak of ERP amplitudes close to the time of response. On visualising the data, it appeared that this peak may be highest during AV trials compared to either V or A trials, which were relatively comparable.

A Kolmogorov-Smirnov test found that the data were normally distributed (all conditions $p > .05$). There were no significant outliers within any of the conditions, and no significant levels of skewness or kurtosis. The data therefore met the initial assumption of being normally distributed to qualify for parametric statistical tests, and we chose to proceed with using a repeated-measures ANOVA as a suitable assessment of any significant main effects. Mauchly's test for sphericity returned as significant and so the assumption of sphericity was violated ($\chi^2(2) = 13.99$, $p = .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .74$). The results found a significant main effect of sensory modality on the peak amplitude of response-locked ERPs ($F(1.48, 48.74) = 4.71$, $p = .022$). This suggested that sensory modality had a significant effect on the peak amplitude of the slopes identified leading up to the time of response. We followed this up with pairwise comparisons using a Bonferroni correction, meaning that all effects are reported at a .0167 level of significance. We found that AV trial peaks ($M = 24.76\mu V/m^2$) were significantly higher than those of V trial peaks ($M =$

$22.07\mu\text{V}/\text{m}^2$; $p = .004$). There was also a trend towards significance that did not pass the Bonferroni correction, suggesting that AV trial peaks were also higher in general than A trial peaks ($M = 22.03\mu\text{V}/\text{m}^2$, $p = .023$).

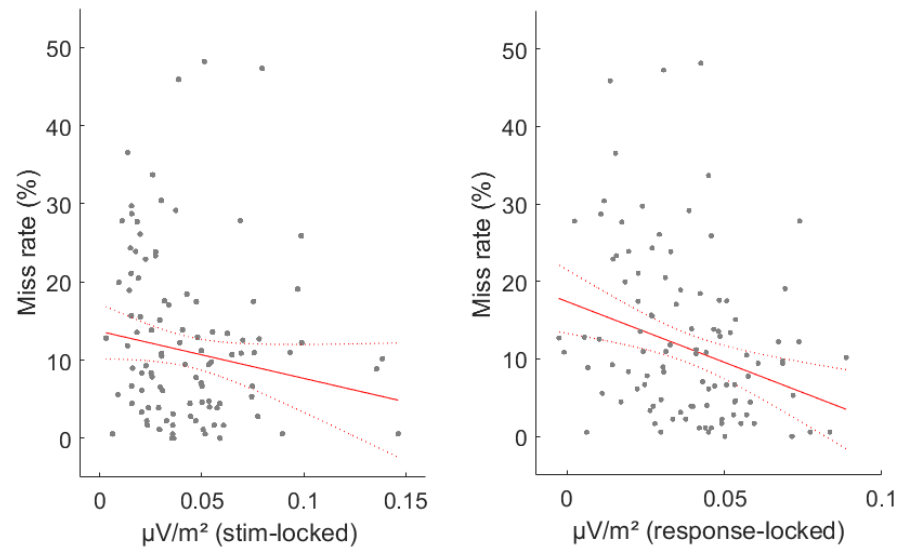
In summary, these results suggest that the peak of ERPs shortly before the time of response were significantly higher during trials with both visual and auditory motion evidence presented, compared to when visual motion alone is present, and that there may be a trend in a similar direction when comparing to trials with sound motion alone. These results may link in with work examining the effect that modulating the amount or quality of motion evidence has on the amplitude of activity observed. In Kelly & O'Connell's original results where motion coherence was modulated over four levels, the amplitude of responses appeared to be higher overall as motion coherence increased, however this effect disappeared when ERPs were centred to the point of decision, as in the second half of our statistical analysis.

Effect of confidence. One possible reason behind differences in the peak amplitudes of our ERPs may be that the presence of additional sensory evidence for motion infers further confidence in the decision made by the participant. More specifically, by presenting coherent motion in both visual and auditory formats, simultaneously and with complementary information, the observing participant objectively has more information at their disposal to make a decision on the direction of motion (or presence of coherent motion), which may make them feel more confident as they make this decision and respond. Previous work has demonstrated increased response amplitudes during decisions which participants indicate were made with more confidence (Gherman and Philiastides, 2015; Philiastides, Heekeren, and Sajda, 2014) . However, if true in our case the difference does not seem to be so apparent as in these works, with smaller overall differences in amplitude.

In order to ascertain whether differences in decision confidence could have been influencing the amplitude of mean ERPs, we conducted a single-trial correlation between peak ERP amplitudes and miss rates. Specifically, we calculated the mean amplitude per single trial of a 50ms window centred on the peak, or the peak component amplitude, and correlated this with a single-trial proxy of

confidence ($1/\sqrt{\text{response time}}$) on a per-subject basis using individual robust linear regressions. This was calculated across conditions, meaning all trials were included. We then assessed the significance of any effect using a one-sample t-test with the resulting β_2 values. This produced a non-significant result ($p = .373$) suggesting that there was no significant relationship between the confidence proxy used and the miss rate of subjects in this task.

A



B

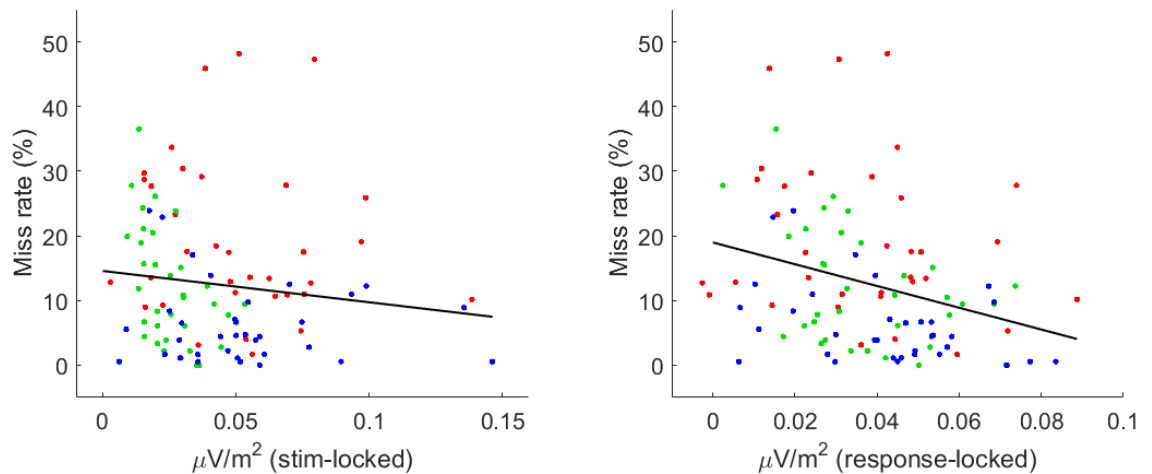


Figure 3.6 Robust linear regression predicting miss rate with ERP slope steepness (i.e. evidence accumulation rate). Stimulus-locked data are visualised in the left-hand figures, and response-locked data in the right-hand figures. **A** Robust linear regression output with single-subject scatter (grey) least squares line fit (solid red) and 95% confidence bounds (dashed red). **B** Repeat of the same distribution of data, with least squares fit line, and with colours indicating sensory condition of each data point (red = visual; green = auditory; blue = audiovisual).

Link to behaviour. While the exploration of our data up until this point revealed some interesting features of the behavioural responses of participants, and the neural activity occurring while they completed the task, this does not in itself suggest any relationship between the two. Establishing a link between ‘brain and behaviour’ is important as this cannot be assumed; these two features of our data could be entirely independent processes and responses unless we show that a link exists between them. In our analysis, we interpreted the steepness of ERP slopes as the rate of evidence accumulation, as in Kelly & O’Connell’s original work. In sequential sampling models of decision making such as the drift diffusion model, the rate at which evidence is sampled affects the overall outcome of the decision, such as whether a decision threshold is surpassed and a choice made and acted upon, as well as the associated speed and accuracy of that decision. We therefore sought to understand whether the rate of evidence accumulation in this task, measured in the steepness of ERP slopes, was related to the overall likelihood that a participant would successfully respond to coherent motion. This would indicate that the effects of coherent motion on the rate of evidence accumulation and the decision to respond were linked, and not necessarily happening due to independent processes.

To explore this effect, we therefore calculated a robust linear regression model, including miss rate as our predictor variable and ERP slope as our response. In other words, we assessed the strength of the relationship between the rate of increase in evidence accumulation observed before the time of response and the likelihood of a given subject being able to detect and respond to coherent audiovisual motion. We chose to use a robust linear regression as it would be less influenced by outliers, and our previous statistical analyses had found that our data were not always normally distributed. Our linear regressions included subject mean miss rates and slopes for all three conditions in the same regression (102 observations total). We computed two robust linear regressions that included either stimulus-locked or response-locked ERP data (Figure 3.6). There was a significant relationship between stimulus-locked ERP slope steepness (i.e. evidence accumulation) and miss rate ($F(2,100) = 5.13, p = .026$), meaning that we were statistically significantly able to predict miss rate using stimulus-locked ERP slopes, however according to adjusted R-squared measures

this only accounted for 3.9% of the explained variability in miss rate, meaning that in practice the model fit could not successfully predict these values. Results for our second robust linear regression were similar; there was a significant relationship between response-locked ERP slope steepness and miss rate ($F(2,100) = 12.2, p < .001$), which accounted for 9.9% of the explained variance according to the adjusted R-squared calculation.

Our results suggest a significant relationship between our neural measure of evidence accumulation and the rate at which subjects could detect coherent audiovisual motion, however further work would be needed to discover which factors account for the variance within this relationship.

To summarise the findings of our analyses: Participants missed significantly fewer target motion periods when coherent visual and auditory motion evidence was simultaneously presented, compared to individually. They were also more likely to respond during trials with higher levels of motion coherence presented. Sensory modality had a significant effect on response times, with participants responding significantly slower during A trials and where less sensory evidence was present (i.e. lower coherence trials). When assessing ERPs following the point of stimulus presentation, A trial slopes were significantly shallower than AV and V trial slopes, suggesting lower rates of evidence accumulation, however when reassessing responses approaching the point of response, we found that AV trial slopes were significantly steeper than V trial but not A trial slopes. We were able to find evidence of a link between ERP slopes and behaviour, but our resulting linear model did not account for a lot of the explained variance. Finally, there were significant differences between the peak amplitude of ERPs, shortly before the time of response, with AV trial ERPs peaking significantly higher than during V trials. Although this suggested confidence may be influencing decision making, a single-trial analysis comparing a confidence-proxy to miss rates found no significant relationship.

3.5 Discussion

In this chapter, we explored how the temporal profile of centroparietal EEG activity during perceptual decision making was altered when participants were able to recruit both visual and (complementary) auditory information. Specifically, we employed a modified version of the classic RDK motion discrimination task (Newsome, Britten and Movshon, 1989; Britten *et al.*, 1992) first described by Kelly and O’Connell (2013), in which participants continuously monitored for coherent motion periods amongst incoherent motion ITIs. This meant that visually evoked potentials typically induced shortly after coherent motion onset were avoided, allowing us to investigate changes in evidence accumulation throughout the trial. We then analysed the behavioural benefit of integrating audiovisual evidence and, in an ERP-centred analysis, compared changes in the rate of ramping activity leading up to the point of response, as well as the link between neural activity encoding evidence accumulation and behaviour. We also explored whether changes in peak activity shortly before responses were related to changes in decision confidence.

As expected, we found that participants were more likely to respond during coherent motion periods when motion coherence was higher, meaning more evidence of visual and/or auditory motion was available. This aligns well with previous literature demonstrating improved decision accuracy with increased sensory evidence (Philiastides and Sajda, 2006b; Philiastides, Ratcliff and Sajda, 2006; Ratcliff, Philiastides and Sajda, 2009; O’Connell, Dockree and Kelly, 2012; Diaz, Queirazza and Philiastides, 2017). Interestingly, participants missed coherent motion periods during similar proportions of A and V trials, but were significantly more likely to respond when audiovisual motion was present. This suggested that the additional evidence presented during AV trials was of benefit as participants made decisions quickly, leading to more accurate decision making. This effect is also well documented in previous research (Gleiss and Kayser, 2012; Raposo *et al.*, 2012; Sheppard, Raposo and Churchland, 2013), and the effect was true whether comparing V or A trials to AV trials. This was reassuring in that we had manipulated the task to a level at which this benefit was relatively balanced, and that one type of sensory evidence was not outweighing the other. However, this balanced improvement was expected as we had selected coherence levels for V and A trial stimuli corresponding with

preselected miss rates during piloting, although we did not do so for AV trial stimuli or take neural responses into account while making this selection.

Our investigation of changes in response time also revealed some interesting effects. Participants responded faster with increasing motion coherence, within all sensory conditions, as expected based on previous research demonstrating this effect (Gold and Shadlen, 2007; Ding and Gold, 2012; Reinagel, 2013). When looking at the pattern of response times across conditions, we saw that participants were considerably slower to respond during A trials compared to both V and AV trials, which were statistically insignificant from each other. This effect raised several questions. As participants were able to detect motion in a comparable proportion of trials during A and V trials, we might assume that the quality of evidence contained in the target motion period was also roughly the same; if this were the case, would this suggest that perceptual decision making involving auditory stimuli are just slower?

Another study with a similarly designed task to our own, although presenting visual or auditory motion exclusively rather than in tandem, found responses to auditory stimuli overall slower (Mulder *et al.*, 2013). However, other historical papers have documented auditory response times as generally faster than visual (Welford, Brebner and Kirby, 1980; Green and Von Gierke, 1984). Another question might be whether the auditory stimulus we used, intended as an equivalent to V trial RDK motion discrimination, might have been an effective match in terms of accuracy but not speed. Whereas V trial stimuli consisted of multiple individual instances of motion, together forming an overall perception of motion in a cloud of moving dots, A trials featured a moving sound that was in essence a singular sound object moving in one direction. Participants might have simply needed to attend to the stimuli for longer to perceive the direction of sound motion, whereas in the same period of time participants would have been able to perceive multiple examples of visual motion direction already, reducing the time needed to make that decision. In the task designed by Mulder *et al.* (2013), the auditory stimulus is comprised of multiple tones that, over the course of the trial, either increase or decrease in pitch. In this way, subjects are able to sample from multiple sounds over the course of the trial, rather than listening for more gradual changes as in the current experiment, so a design

similar to this may be more appropriate. However, their task does still require participants to monitor for a longer period of time over which a general direction of pitch can be perceived. For this reason, perhaps a follow-up study could establish a more appropriate sound-equivalent of the RDK stimulus, for use in an audiovisual decision making task.

We investigated how electrophysiological responses varied between audiovisual and unisensory (A or V only) trials by using an ERP-centred approach, and averaged across participants and coherences both when centring on the onset of coherent motion (stimulus-locked) and on the time of response (i.e. a button press; response-locked; see **EEG data analysis**). We also selected sensors that, during piloting and from visual inspection of ERP traces averaged across conditions, clearly demonstrated gradual increases leading up to the time of response (see Figure 3.3). Stimulus-locked ERPs revealed that AV and V trial slopes were significantly steeper than A trial, implying faster rates of sequential sampling and therefore evidence accumulation. This was clear from the visibly delayed peak in activity during A trials compared to V/AV, which corresponded with the slower responses during A trials, and motivated our decision to transition to a response-locked view of ERPs; in theory this would allow us to more clearly compare evidence accumulation rates between trial types, and indeed this appeared to be the case once the corresponding figures had been generated (Figure 3.5). This view did seem to provide a more accurate temporal profile of activity leading up to the time of response, once the latency differences had been accounted for. In this view, we discovered that AV trial slopes were significantly steeper than V trials, but not A trials (although there was a non-significant trend in this direction). Indeed, A and V trial slopes were comparable which mirrored the similarity in miss rate found during these two conditions. Although we cannot factor in AV versus A trial slopes, the results suggested that audiovisual perceptual decision making was enhanced, compared to visual only, both in participants' ability to detect coherent motion and in the rate at which they accumulated sensory evidence. This is consistent with previous research demonstrating positive changes in neurological activity during trials with complementary audiovisual stimuli (Bernstein, Auer and Takayanagi, 2004; Lippert, Logothetis and Kayser, 2007; Bernstein *et al.*, 2008; Raposo *et al.*, 2012; Mercier and Cappe, 2020). However, our attempt to link neural and

behavioural improvements was only moderately successful; while a robust linear regression was able to find a significant relationship between the two, the linear model that we fit did not explain much of the variance seen. We would therefore recommend that future work should investigate this relationship more closely and include other factors that might have contributed to this relationship, such as coherence, sensory condition, and confidence.

Following on from our analysis of evidence accumulation changes, we sought to explore differences in peak activity shortly before the response, specifically to examine whether the presence of complementary auditory evidence might lead to increased peak activity compared to either unisensory condition. Our analysis found that participant ERPs peaked at significantly higher levels during AV trials compared to V trials, however there was no significant difference between AV and A trials. One possible interpretation of this result, from the perspective of sequential sampling modeling, is that in the presence of more sensory evidence the threshold for a decision may be higher. When both visual and auditory evidence is present, there is also a greater amount of noise, and a higher decision threshold accounts for this by requiring more evidence to accumulate (therefore assuming that some may be noise). This may be why we saw higher ERP peaks just before responses, however this effect was not as pronounced as in other work where the same modality in varying coherence levels is used. We have captured this data in our work for potential further investigation, however as our audiovisual condition only used one auditory coherence level it may be necessary to repeat the experiment with further coherences included.

Another interpretation of varied pre-response ERP peaks could be varying levels of decision confidence. Other research employing unisensory perceptual decision tasks have linked self-reported levels of decision confidence with changes in pre-response amplitude (Kiani and Shadlen, 2009; Ding and Gold, 2013; Gherman and Philastides, 2015, 2018). We investigated whether the increased availability of sensory evidence created by the inclusion of helpful information from more than one modality may have increased the levels of confidence participants had when making their decision. If this were true, we should see a positive relationship between a measure of confidence (here a proxy measure using single-trial response times) and subject-specific pre-response peak amplitudes, meaning

that trials where participants were potentially more confident were also trials with higher peak amplitudes just before they responded. Our analysis did not find any such relationship, however it was not exhaustive in nature and there remains potential to investigate this effect. One limitation of this exploration was that we did not incorporate a direct measure of confidence into the task, meaning that we could only estimate what levels of confidence were on each trial and not use data explicitly reported by the participants themselves. Further research could try to incorporate this into the task, however this may come at the cost of the gradual nature of the task; if a visual prompt to report confidence were included after each trial, that would undo the otherwise seamless transition from ITI to target motion period, and participants would have a clear signal that a motion period had just finished. However, if a successful incorporation of measuring confidence can be achieved, researchers could investigate the interaction between audiovisual motion evidence (versus unisensory) and confidence on decision processing, in a task that allows a clear observation of evidence accumulation.

Other limitations of this task may exist outside of effectively capturing confidence. As we reported in our results, once trials where participants failed to respond were removed, participants could discriminate the direction of motion at a high level of accuracy (>97% in all conditions within the current dataset). Based on observing this effect during piloting, our decision moving forward to the full study was to use miss rate as our behavioural measure, as this did scale with coherence and was not obscured by a ceiling effect. Indeed, Kelly & O'Connell report high levels of accuracy and used miss rate as a primary behavioural measure in their original study. This raises significant questions as to the type of decision making being captured by this task; if, when participants do respond, they are able to discriminate motion direction in a large proportion of trials, it may imply that a significant amount of the task difficulty is in successfully detecting coherent motion and not solely in discriminating its direction. This does not mean that the decision making process is not being captured in our EEG data, only that a considerable detection component may be being included in the activity we see. The temporal profile of our results certainly appears to follow the gradual ramping up in activity documented by other perceptual decision making research not using this task (Philiastides and

Sajda, 2006b; Philiastides, Ratcliff and Sajda, 2006; Ratcliff, Philiastides and Sajda, 2009), suggesting that perceptual decision making is a key component of the signal recorded in this task, however the recorded behavioural results indicate a high demand on detection processing. This may be a factor in why our attempt to correlate confidence with pre-response amplitudes was unsuccessful; changes in detection processing before discrimination will have impacted on non-decision time and therefore the overall response times of each trial, and as our confidence proxy used response times this may have clouded any effect.

One proposed solution to the overweighted role of motion detection in this task may be to include a cue prior to the start of a target motion period, however the timing and presentation of said cue would have to be careful and avoid creating the VEPs that the task was indeed designed to prevent. An auditory cue presented roughly a second before the onset of coherent motion would avoid creating VEPs and may be recent enough to direct additional attention to the task, aiding in motion detection but still requiring motion discrimination to occur. This may also mediate the issue of drifting attention during the task; during piloting, participants indicated feeling tired of the task relatively quickly, and this often seemed to be due to the lack of any clear stimulus changes during the task. While this was by design, we factored this effect into the experiment design and made trial blocks as short as possible. Though not examined empirically, clear alpha waves were visible in the raw EEG data as participants completed the task.

Our previous results chapter (see 2.4 Results) found changes in late post-sensory processing during audiovisual decision making, compared to during visual decision making, and in opposition to changes in early sensory processing. This follows a debate on whether the integration of multisensory evidence for decision making occurs earlier on, shortly following stimulus onset and while Here, with the effects of VEPs significantly reduced if not eliminated, we did not observe many early changes in processing. Activity began to diverge from roughly equivalent amplitudes shortly after stimulus presentation, but became distinct in the rate at which activity ramped up as the point of response approached. This was true whether looking at the data from a stimulus- or response-locked view. One point of interest was that A trial activity seemed to

ramp up later into the trial (~100ms later or so) compared to V or AV trial activity. Indeed, this was our motivation behind fitting linear models to A trial data using a later start time than for the other two conditions, in order to best capture the rate of interest. However, this may be due to factors induced by the task design such as changes in non-decision time and detection-related activity, as we have already discussed.

Regarding the selection of analysis windows for stimulus-locked and response-locked data, per condition, which were based on visual inspection; if this experiment were to be repeated, these windows would ideally have been selected using methodology less prone to bias. For example, one approach could be to select the start and end points of the windows using definitions set out before seeing any data, or computationally, for example, by attempting to fit flat and sloped linear regressions to see where modelled data began to increase over time.

That we found increased rates of evidence accumulation during audiovisual decisions, compared to visual, also aligns with the findings of our previous chapter. Specifically, in the previous chapter we used a neurally-informed cognitive model and found that the improvements in behavioural decision making observed (i.e. that participants were more accurate when discriminating audiovisual stimuli versus visual) could be explained largely by increased rates of evidence accumulation. It is encouraging that when participants completed this task involving ‘lower complexity’ stimuli, and with a different approach to measuring changes to evidence accumulation rates, we continued to see the same effect. This overall gives credence to the understanding that key components involving perceptual decision making, and the integration of audiovisual evidence for that decision, occur later in the trial (closer to the time of decision making).

Whereas in our previous chapter we were able to directly predict behavioural improvements using neural data, our attempt to do so with this task and our ERP focus was less successful. The most significant difference is that the relationship found in the previous chapter used single trial data as part of a more rigorous neurally-informed hierarchical drift diffusion model, whereas the current chapter used participant-averaged data in a relatively simple robust linear

regression. The variability in rates of accumulation in each trial could contain useful data, as sequential sampling models understand decision making to involve the noisy accumulation of evidence, and by averaging across trials in the current chapter this information is lost. We also did not include other factors in our model, therefore a follow up analysis for the current chapter may involve using single trial fits of accumulation rates (i.e. slopes) to predict response time or miss rate, ideally including sensory modality and coherence as contributory factors.

In summary, the use of this continuous version of the RDK motion discrimination task is a useful tool to examine changes in evidence accumulation rate, and supports the findings of our previous chapter in that providing audiovisual evidence of a decision, as opposed to visual alone, enhances decision performance. Furthermore, this seems to be linked to improvements in the rate of evidence accumulation. However, there were several limitations to this work that may be addressed by a) adjusting the task design to prevent detection featuring too predominantly in the neural activity captured, which might also allow us to compare true decision performance (accuracy) without a ceiling effect, and b) by incorporating some of the analytical techniques used in the previous chapter, such as using single-trial data in our analysis and perhaps including it in a neurally-informed cognitive model of decision making as we did successfully in Chapter 2.

4 Temporal characterisation of oscillatory activity during audiovisual perceptual decision making

4.1 Summary

The thesis so far has supported theories that the integration of evidence during multisensory decision processing (here audiovisual specifically) is reflected in enhancements to evidence accumulation later on in the course of a decision. Additional auditory information appears to boost behaviour performance, and this enhancement is linked to augmentations of late, post-sensory decision processing, as opposed to early activity. Our analyses thus far focussed on electrophysiological information captured by EEG, however this information has been used to demonstrate oscillatory patterns in neural activity that can reveal further information on the neural correlates of perceptual decision making. Here, we performed a reanalysis of data collected in Chapter 3, instead using a spectral analysis to decompose the broadband signal recorded into its component parts. We again discovered a gradual change in power leading up to the time of perceptual decisions in a pattern mirroring evidence accumulation processing, however the differences in evidence accumulation rates did not reveal the same enhancements specifically during trials with audiovisual sensory information, versus visual or auditory only, in the same way that we had seen in our previous chapters. Some changes in oscillatory activity indicated a potential role of premotor areas in evidence accumulation, as suggested by the embodiment hypothesis. Further work would investigate this possibility more directly.

4.2 Introduction

In the previous chapter, we explored changes in broadband electrophysiological neural activity relating to the process of evidence accumulation (EA), when making an audiovisual perceptual decision. Specifically, we explored how, approaching the point of response, EA is modulated by the type and strength of sensory evidence presented (V, A, or AV). In this experiment, participants found it easier to perceive and respond to AV motion compared to either unisensory variant (V or A). Following a response-locked ERP analysis, we observed a steeper ramping up of EA activity in the period approaching the decision, specifically as in the CPP component previously identified by O'Connell *et al.* (2012).

While this analysis was successfully employed to identify temporal electrocortical signals linked to the process of perceptual decision making (PDM), by its nature it was only able to provide a direct view of the ramping up of broadband activity approaching the response, although our analysis may have extracted activity related to specific frequency bands. Often, complementary information to that of broadband neural activity can be discovered by using a spectral analysis to decompose the signal into its component frequency bands; a typical event-locked broadband signal tends to be dominated by lower frequency activity, such as that of delta (1-4Hz), theta (4-8Hz), and alpha (8-12Hz) frequencies, and as a consequence any activity reflected within higher oscillatory patterns such as beta (13-30Hz), (low) gamma (30-70Hz), or high gamma (70-100Hz) can be obscured. The overrepresentation of alpha band activity could have caused some issue if relying only on the broadband analysis, as the design of our task required a continuous monitoring of dot motion stimuli, with no cues as to the start or end of trials. This may have kept participants more engaged or alert due to the need to continuously and actively search for coherent motion. On the other hand, some may have felt disengaged due to the lack of cues or obvious stimulus changes during each block. Importantly, lack of attention has been linked to increased alpha activity (Cooper *et al.*, 2003; Herring *et al.*, 2015), which may have obscured information from within other frequency bands when taking a broadband approach to our investigation.

For this reason, we were motivated to conduct a time-frequency analysis, as there is potential to reveal additional complementary information about the PDM process unfolding during our audiovisual task. By decomposing the broadband signal, we were able to observe whether there were changes to the rate of EA or the setting of decision boundaries, and whether those were seen specifically within certain frequency bands. Any effects on the rate of EA would be apparent in the speed with which activity began to ramp up as the time of decision approached (FitzGerald *et al.*, 2015) whereas changes to the setting of a decision boundary would be seen in a general modulation in the magnitude of that activity at the time of, or shortly before the time of, the response (Philiastides, Heekeren and Sajda, 2014; Gherman and Philiastides, 2015). In investigating these effects, and comparing the activity revealed in each of our sensory conditions, we would be able to observe whether the presence of complementary audiovisual information during our motion discrimination task had specific effects on EA or the setting of a decision boundary during said task. As participants were presented with unisensory trials of each sensory modality (V and A) as well as the AV trials, we were also able to observe whether those changes were linked more to the addition of one type of sensory evidence over the other; that is, if more change in activity was seen in AV trials compared to V trials or A trials alone.

Following insight into broadband analyses of EA and decision boundary effects, the use of a time-frequency analysis on PDM activity has been a focus of some previous research in the field. Several studies have investigated whether specific frequencies of oscillatory activity have individual roles within these PDM processes. The coordinated patterns of excitatory and inhibitory activity seen during perceptual decisions, and that contribute to the ramping of ERPs seen in the previous chapter (O'Connell, Dockree and Kelly, 2012), have been observed to lead to gamma-band oscillations (Bollimunta and Ditterich, 2012; Buzsáki and Wang, 2012). Further, others have found monotonic increases in gamma activity within EA areas (Wang, 2002), suggested in Polanía *et al.* (2014) to be reflective of large collections of neurons in coordinated activity detectable via the readout of extracellular electric fields (Buzsáki and Wang, 2012).

The 2014 study by Polanía considered the potential neural links between value-based and perceptual decision making, specifically whether overlap exists between the two in terms of the neural computations involved that seem to accumulate evidence in favour of one choice or another. To do so, they recorded EEG data during a task where trials differed only in the type of decision required (preference or perception), while the stimuli and general design remained the same. They fit a simple sequential sampling model (SSM) to their behavioural results, using this to make trial-specific predictions of EA signals with the EEG data, which had been decomposed using a time-frequency analysis. Specifically during PDM trials, these predicted EA signals were found in gamma oscillations (48-66Hz) within parietal sensors, however in contrast to their findings regarding value-based decision making, they did not find any predicted activity from sensors located over frontal regions. Further, they found that lateralised readiness potentials (LRPs) within motor areas were highly correlated with the SSM's predictions during perceptual decision trials (as well as value-based trials), interestingly in areas very similar to the CPP cluster identified during PDM tasks.

This study suggests key roles of parietal gamma-band activity and motor-related LRPs in PDM, in that they reflect and seem to embody EA as the time of decision is approached, rather than simply driving the motor response that instigates the decision. This is consistent with the relatively recent embodiment hypothesis (Rorie and Newsome, 2005; Filimon *et al.*, 2013); whereas previous work supported the role of areas such as the dorsolateral prefrontal cortex as an abstract decision module, with motor and premotor areas primarily being involved in carrying out the outcome of a decision (i.e. a button press), the embodiment hypothesis instead suggests that these areas may play a role in the formation of the decision itself (Tosoni *et al.*, 2008). Indeed, a recent animal multi-neuron recordings paper revealed that, while decision representations in sensory and association cortex were sufficient to perform the task, inactivation of a downstream premotor area led to gross behavioural impairment (Wu *et al.*, 2020).

Further evidence regarding beta-band activity was highlighted in work by Donner *et al.* (2009), in addition to similar findings regarding gamma-band and sensorimotor decision activity. The experimenters recorded MEG activity while

participants completed a “yes/no” visual motion task, responding when they identified coherent motion within a random dot kinematogram (RDK) stimulus, and with trials labelled according to signal detection theory (Green and Swets, 1966). Using a multitaper spectral analysis (Mitra, 2007) and adaptive spatial filtering, they identified gamma- and beta-band activity that predicted the choice of participants seconds before the choice itself was made by a button press. This activity gradually increased towards both possible choices during stimulus presentation, an observation possible due to the lateralisation of each response (yes/no choices were made with a left/right hand button press exclusively). They also found such choice-predictive activity was linked to the temporal integration of gamma-band activity in the motor cortex, suggesting that motor planning for each respective choice was the result of continuous sensory evidence accumulation as participants viewed the RDK stimuli. Other studies have found similar patterns of increasing beta-band activity prior to perceptual decision (Siegel, Engel and Donner, 2011; O’Connell, Dockree and Kelly, 2012; Wyart *et al.*, 2012). Such beta-band activity has been described as reflecting the “temporal and spatial dynamics of the accumulation and processing of evidence in the sensorimotor network leading to the decision outcome” (Haegens *et al.*, 2011). These findings again support the embodiment hypothesis; rather than simply being fed instructions to “push the button” from a separate decision module, sensorimotor areas themselves may contribute to the decision to do so, i.e., the motor decision is embodied within motor and premotor regions.

After seeing clear changes to ramping activity leading up to the point of response in our previous analysis, we decided to investigate whether any further frequency-band-specific effects had been missed by using a broadband approach, with particular regard to the effects of AV integration on perceptual decision making. By reanalysing the electrophysiological data previously described in Chapter 3 (see 3.3.4), and decomposing it using a time-frequency analysis, we were able to investigate whether the integration of AV evidence elicited particular changes to decision making activity within frequency bands such as beta or gamma. Specifically, we would investigate changes to oscillatory activity linked to decision variables such as ramping activity (EA) or the peak of activity shortly before the response (a decision boundary).

In order to best frame our approach, we mirrored the analysis of our third chapter. Assuming the centroparietal cluster reflects the main generator of the relevant decision activity (O’Connell, Dockree and Kelly, 2012; Kelly and O’Connell, 2013; Polanía *et al.*, 2014), we decided to decompose it further into its constituent frequency bands. We focused our analysis on the centroparietal positivity component (CPP) highlighted in our ERP analysis and previously identified by O’Connell *et al.* (2012), as this region seemed to reflect the main generator of the relevant decision activity to this task. Further, we decided to decompose it into its constituent frequency bands with a time-frequency analysis, and investigated whether we would see similar changes to activity near the time of the response, here in the form of a ramping up of the power of oscillatory activity or in a change to the peak of decision activity, the latter of which would suggest a change to the decision boundary. To observe these effects, we followed the same method of statistical analysis as in the previous chapter, instead comparing the power estimates across time and between conditions. In doing so, we would be able to clearly compare and translate any effects seen between the broadband and time-frequency analyses. In particular, following on from the work of Polanía *et al.* (2014) and Donner *et al.* (2009), we were interested in changes to beta- and gamma-band power leading up to the time of response, and whether the additional sensory evidence provided during AV trials would impact on features of perceptual decision making such as evidence accumulation or the setting of a decision boundary. Due to the additional benefit that AV trials should provide in terms of the strength of sensory evidence, one finding we expected to see was on EA, specifically in an increased rate of ramping up of beta- or gamma-band power approaching the time of decision, with AV trials having a steeper ramp compared to A or V trials alone. This would be reflective of a faster rate of EA or increased drift rate, potentially due to the additional sensory evidence provided in this condition. Further, we expected we might see a modulation of the peak beta- or gamma-band power shortly before the response, with AV trials having a higher peak than A or V trials. This difference would reflect a change in the threshold of sensory evidence set to reach a decision.

4.3 Materials & Methods

This study is based on reanalysis of data presented in Chapter 3 (see Materials & Methods). All methodological details relating to participants, stimuli and behavioural paradigm, as well as EEG data acquisition and pre-processing, are identical unless otherwise specified.

4.3.1 Participants

Thirty-four paid participants (20 female, age range 19-33 years) who took part in the experiment were included in the full analysis; thirty-nine completed the task however five were discounted due to data collection issues ($n=1$) and poor-quality data recording ($n=4$). All were right-handed, reported normal or corrected to normal vision, normal hearing, and no history of neurological problems. The study was approved by the college of Science and Engineering Ethics Committee at the University of Glasgow (CSE 300150102) and informed consent was obtained from all participants.

4.3.2 Stimuli and task

Stimuli and the behavioural paradigm are described in more detail in Chapter 3 (3.3.2). In short, participants were asked to attend to a continuous version of the classic RDK motion detection task (Kelly and O'Connell, 2013) and discriminate motion direction without clear indication as to the start of coherent motion periods (target periods). Additionally, we altered the task to include a simultaneously-presented continuous auditory stimuli where white noise would move to the left or right in parallel with coherent RDK motion. In the place of a normal 'blank' inter-trial interval (ITI), we presented either incoherently moving dots or a static sound, meaning the types of stimuli used in the target period were present but did not contain any useful sensory information regarding a direction of motion. By presenting an almost undetectable transition between the ITI and target period, we aimed to prevent causing ERPs due to the sudden change in stimulus presentation. Instead, we changed the information contained

within it, which allowed us a relatively uninterrupted view of the electrophysiological features of the decision making process over time. With the task split into 15 blocks, each block presented coherent motion visually (V trials), aurally (A trials), or audiovisually (AV trials) exclusively. Trials also varied within each block between two motion coherences, therefore varying the difficulty of the task within each sensory condition.

Participants were asked to fixate on a central cross and were informed via on-screen text of the stimulus type that would contain coherent motion five seconds before beginning (i.e. V, A, or AV trials), but were requested to monitor all stimuli throughout as while piloting this task we found that some participants chose to close their eyes during A trials. They were instructed that, in the instance they perceived coherent motion in either a leftward or rightward direction, they should indicate the direction as quickly as possible with a button press. During each trial, following an incoherent motion interval of 2s, 3.5s, or 5s, we presented coherent motion for 1.9s. This was a fixed amount of time regardless of if participants responded before the end of the target period. The target period would then transition into the next ITI and then another target period, without an obvious cue or indication as to the boundary between the end of the ITI and the start of the target period.

Prior to the full task as described above, participants completed two shorter training tasks with lowered difficulty. This was to ensure participants had a basic level of familiarity with the task design and stimuli before moving on to the extended full task. Task difficulty was controlled using, for dot and sound stimuli respectively, the proportion of dots moving together coherently or the ‘distance’ by which the sound stimulus travelled from the centre to the left or right (see 3.3.2 for further explanation).

4.3.3 EEG data acquisition

Continuous EEG data were recorded using a 64 channel EEG amplifier system (BrainAmps MR-Plus, Brain Products, Germany) with Ag/AgCl scalp electrodes placed according to the international 10-20 system on an EasyCap (Brain

products GmbH). Channels were referenced to the left mastoid, with a chin electrode acting as ground. Data were sampled at a rate of 1000 Hz at an analogue band pass of 0.0016-250Hz. To obtain accurate event onset times, experimental event codes and participant responses were recorded simultaneously with the EEG data using PsychoPy and Brain Vision Recorder (Version 1.10, Brain Products, Germany).

4.3.4 EEG data pre-processing

The data were pre-processed offline (excluding the previously mentioned analogue band pass, see EEG data acquisition) using MATLAB (version 2015b, The MathWorks, 2015). We applied a 0.5-100Hz bandpass filter to remove slow DC drift, retaining higher frequency data for our time-frequency decomposition later in the analysis. We also used data recorded in an eye-movement calibration task to identify linear artefacts associated with eye blinks and eye-movement, using a principal component analysis approach as described in Parra *et al.* (2005) to remove them. The data were re-referenced to the average of all channels. Following eye-movement removal, further trials were removed if participants responded so soon after the onset of coherent motion that it was unlikely that they were responding to stimulus changes (i.e., trials where the response time was faster than 300ms post-stimulus), where amplitudes exceeded 150 μ V, or where participants failed to respond within the target period were removed.

4.3.5 EEG spectral analysis

The spectral analysis and code were adapted from Gherman (2017), including custom code which was edited and restructured for the present work. We performed two spectral analyses using the FieldTrip toolbox (Oostenveld *et al.*, 2011) and custom MATLAB code. These analyses varied only in the tapering method used, with all other parameters consistent between the two. Pre-processed data were segmented into epochs from -1750ms to +1000ms relative to the time of response. Each time-frequency decomposition was performed

separately per channel and per trial, with windows centred from -1250ms to +500ms in 50ms steps, again relative to the time of response. We computed time-frequency representations of the EEG signal at 50 frequencies (2-100Hz, in steps of 2Hz), using a sliding-window Fourier transform. As in our analysis in Chapter 3, we focused on the same six electrodes (CP_Z, P₁, P₂, PO₁, PO₂, O_Z) roughly positioned over the centroparietal positivity (CPP) component identified by Kelly and O'Connell (2013), however we ran the full spectral analysis as described here on data from all 64 electrodes.

Before the Fourier transform, we multiplied our windows of interest with, separately per analysis, a) a single Hanning taper or b) using the Multitaper method. We chose this approach as, while the Hanning taper is commonly employed and useful for high signal-to-noise ratio situations (i.e. beta-band activity), at higher frequencies where the signal-to-noise ratio can be lower (i.e. gamma-band activity) a Multitaper method may be preferable due to its ability to improve the signal-to-noise ratio (Cohen, 2014). The latter should not, however, be used when the frequencies of interest lie within the beta-band as the otherwise useful spectral smoothing feature can impede frequency isolation in this range. Therefore, by using two separate tapering methods we were able to optimise our analyses for our frequency bands of interest; beta (16-30Hz), gamma (30-64Hz excluding 50Hz to account for AC power noise), and high gamma (64-100Hz). We used the power estimates of one tapering method exclusively per frequency band, specifically those of the Hanning taper method of beta-band power, and the Multitaper method for both gamma- and high-gamma-band power.

For both analyses, we also aimed to find an ideal balance between the spectral and temporal resolution of our power estimates (Cohen, 2014) by adapting the length of the sliding window per frequency. To do so, we modulated the number of cycles per frequency, using logarithmically-spaced numbers of cycles rounded to the nearest integer, with the number of cycles ranging from 4 for the lowest frequency (1000ms, 2Hz) to 16 cycles for the highest frequency (160ms, 100Hz). Specifically for the Multitaper spectral analysis, we computed these representations with frequency smoothing set to scale with frequency, three

orthogonal Slepian tapers, and the same method of adjusting the sliding window length.

Following the time-frequency decomposition, the resulting single-trial power estimates were then averaged across trials within each sensory condition (A, V, and AV) and each subject. These values were baseline-normalised relative to the power within the period -500ms to -200ms pre-stimulus onset, calculated using sliding windows in steps of 50ms for a total of 5 windows, which varied in length per frequency as in our main analyses. To do so we ran two other spectral analyses locked to the time of stimulus presentation, stored the baseline value from these data, and applied them during the baseline correction of the response-locked spectral analyses described here. All other parameters were kept the same while doing so, meaning that the data were comparable in all ways except timing. We averaged the power estimates across all conditions, separately per subject and frequency, and by averaging across conditions we could increase the signal-to-noise ratio of the baseline calculation that followed. This correction was applied to all condition-averaged power estimates on a subject- and frequency-specific basis. From here, these data were used in further statistical analyses.

4.3.6 Statistical analysis

Much of the structure of our statistical analysis was based on that of Chapter 3 (see 3.3.6), with a focus on identifying periods of peak activity, and the ramping up or down of said activity, approaching the time of response. In the place of ERP data and amplitudes we harnessed power, with the spectral analysis having produced power estimates for each of our three conditions (V, A, and AV). Throughout our statistical analysis, we separated these power estimates into three separate frequency bands; beta-band (16-30Hz), gamma-band (30-64Hz, excluding 50Hz to avoid noise from AC power supplies) and high gamma-band (64-100Hz). These definitions were based on those of Donner *et al.* (2009).

In order to explore the temporal profile of these power estimates, we first calculated grand-averaged power over time per frequency band. Specifically, we

calculated the mean across conditions, subjects, electrodes from our specified cluster, and frequencies within each band, to produce a grand mean of power over time per band. In doing so, we were able to specify the times of interest for further analysis without the influence of any condition-specific effects. From here, we defined specific time points from which the rate of change in power appeared to increase considerably, and time windows surrounding the point of peak power, producing a 'start' time and an 'end' window per period of interest identified.

Having defined these time periods, we then extracted subject-specific peak frequencies and peak times, separately per condition and frequency band, for each period of interest. To do so, we used the condition-specific power estimates of each subject, still decomposed into separate frequencies in steps of 2Hz, and calculated the minimum or maximum power across frequencies within our defined windows, depending on whether power had been increasing or decreasing prior to this period (i.e. following a period where power was ramping down, we calculated the minimum power within the peak window, and where power had ramped up we calculated the maximum). This produced the specific frequency within each band and per condition of peak power, for each of the time points contained within our defined windows (in 50ms steps). We then calculated the minimum/maximum of these to produce a single peak frequency and time point per subject, per condition. We used this process in full for each period of interest, where power increased or decreased rapidly near to the time of response.

We used these peaks to analyse peak power differences between conditions (V, A, and AV), specifically whether the complementary sensory information provided during AV trials caused power to reach a significantly different level to V or A trials before the time of response or otherwise. To statistically compare the difference in peak power between conditions, we ran one-way repeated-measures analyses of variance (ANOVA) which compared V, A, and AV power across subjects. Before each ANOVA, we computed Mauchly's Test for Sphericity, and corrected the output of the ANOVA where this assumption was violated. Where the ANOVA revealed a significant effect of the condition on peak power, we computed pairwise comparisons to identify specifically which comparison of

conditions was driving the overall effect. In this case, we adjusted for multiple comparisons using the Bonferroni correction. We calculated the statistical difference between the peak power of V, A, and AV trials using this method for each frequency band and period of interest.

In a further analysis, we sought to investigate whether the condition had any effect on the rate of change in power over time, in other words the slope, as we were interested in whether the differences between sensory information available in each of the conditions would lead to changes during the decision process. These changes could be related to the accumulation of evidence, if looking shortly before the time of decision, or potentially in a measure of confidence, if focusing on the time period of or shortly after the response. In one possibility, if the rate of change in power leading to or just before the time of response was greater in AV trials compared to V or A trials, this could suggest that the greater amount of evidence present led to increased rates of evidence accumulation leading up to the point of decision. To explore this, we calculated, on a subject-specific basis, linear fits to the data between the start points we had already defined and the subject-specific peak times we had subsequently extracted. We also calculated these linear fits using the subject-specific peak frequency per band we had identified in the same process. For clarity, for each subject, condition, frequency band, and period of interest, we fit a slope to the peak frequency of each subject between the time at which power began to rapidly increase or decrease and the point at which power peaked shortly afterwards.

Finally, we used the estimate of the slope, m , to compute statistical comparisons between the slope estimates of each condition, across subjects, per frequency band and period of interest. We used the same approach as in the comparison of peak power, computing one-way repeated-measures ANOVAs per frequency band and period of interest, with appropriate corrections and pairwise comparisons where a more detailed breakdown of the statistical difference present was needed.

4.4 Results

4.4.1 Behaviour

All behavioural results are presented in the Results section of Chapter 3 (see 3.4.1). Importantly, we showed that subjects were more likely to perceive coherent motion during AV trials (7.63% missed trials) compared to V or A trials (21.81% and 28.19% of trials respectively; $F(1.44, 47.59) = 12.21, p < .001$), and respond with the direction of perceived coherent motion during the target period. This was despite initial intentions to assess decision accuracy as the main proxy of decision making during the task, however we found that, once missed trials were removed, the average accuracy across the remaining trials was close to ceiling. As we were able to demonstrate differences in behaviour with the type and amount of sensory information presented, and our ERP analysis revealed that there were clear periods of electrophysiological activity up to and at the time of response, we argued that the task did capture a decision making process. Further, our results were consistent with those of Kelly & O'Connell who also reported miss rate as their main behavioural measure of task performance. For this reason, this result is comparable with that of other literature that demonstrates that audiovisually-informed decisions are more accurate than those informed by sound or vision alone (Chen *et al.*, 2011; Raposo *et al.*, 2012; Gleiss and Kayser, 2013, 2014).

4.4.2 Spectral analysis

As the grand-average calculation showed rapid changes in power both approaching and shortly after the response in all frequency bands (Figure 4.1), we decided to focus our analysis on two periods of interest that we hereafter refer to as a pre-response period and a post-response period. The specific temporal definitions of these are defined in Table 4.1.

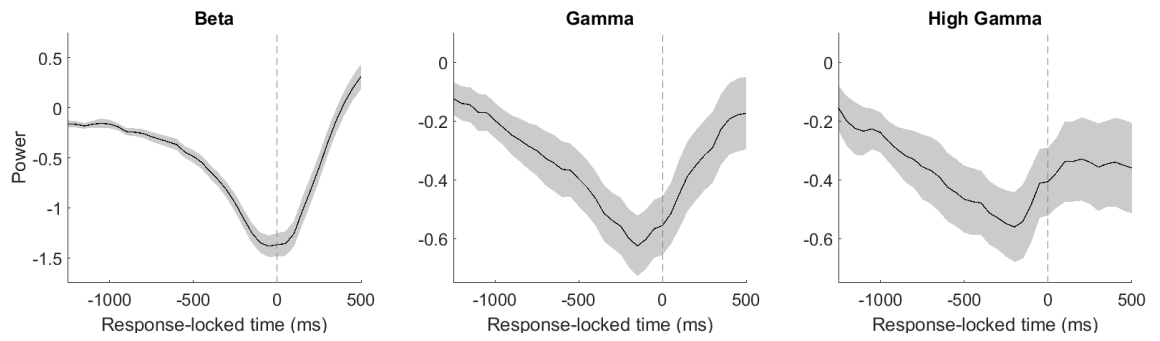


Figure 4.1 Temporal evolution of power estimates, relative to the time of response. Mean calculated across subjects **CCP cluster sensors**, conditions, and frequencies (within each frequency band). Error bars are standard error of the mean (SEM). Dashed line indicates time of response.

Period of interest	Frequency band	Beginning of slope (ms)	Peak selection window (ms)
Pre-response	Beta	-550	-100 : 0
	Gamma	-750	-250 : -50
	High-gamma	-750	-300 : -100
Post-response	Beta	0	+200 : +300
	Gamma	-150	+200 : +300
	High-gamma	-200	+50 : +150

Table 4.1 Specific timings of slope and peak power per period of interest and frequency band. All times are relative to the time of response.

Window definition. Firstly, for the pre-response period, we selected ‘start’ timepoints at which power seemed to begin ramping down more steeply (Figure 4.1), with one timepoint defined per frequency band. For gamma- and high gamma-band power, there was an extended period over which power ramped down quite consistently from close to the start of our data epoch, without a clearly-identifiable beginning, so in this instance our chosen time was based approximately on the average response time of V and AV trials¹ to make sure that we were capturing a period consistently within the decision process on each trial. For beta-band power, this timing was based on the point at which a clear downward slope was evolving, which was also consistent with that which we defined in the ERP analysis of the previous chapter. Next, we defined time windows that encompassed the period of peak power per frequency band. For beta-band power, as clear in Figure 4.1, this was defined as shortly before or at the time of response (Table 4.1). However, for gamma- and high-gamma band power, this peaked before but not at the time of response, resulting in window selections that spanned either side of these peaks but not overlapping with the time off response.

We then defined the timings of the post-response period of notable change in power. In opposition to the pre-response period, power ramped up during this time, and we later used these peak window definitions to identity the time and frequency of maximum power per participant (as opposed to minimum power). Beta-band power clearly ramped up sharply shortly after the time of response and continued to do so until the end of our selected epoch, therefore this period began at the time of response and ‘peaked’ a few tenths of a second later. The peak window was defined with potential subject-specific variation in this change in power in mind, and to make sure we captured the period of clearest power change following the time of response. For gamma-band power, the start of change in power was defined as shortly before the response, with the peak window consistent with that of beta-band power. Finally, high-gamma band power also began ramping up before the time of response, however the peak

¹ The average response time of auditory trials was considerably longer than that of visual and audiovisual trials, and so to prevent capturing periods of time that would precede the point of stimulus presentation in some instances, we based our selection of a start time of ramping down to account for this caveat and ensure that the window more consistently reflected the decision process.

window was defined slightly earlier than for beta or gamma, as it appeared to level off shortly after the response itself.

Using these timings, we extracted subject-specific peak power estimates per condition (V, A, and AV) and frequency band (beta, gamma, and high-gamma), along with the associated timings and frequency of these peaks. We averaged across these subject-specific values to produce an estimate of mean power over time, per condition and frequency band (Figure 4.2). We also specifically used the peak frequencies produced per period of interest's respective definitions, resulting in two versions of the condition-specific group-averaged power estimates per analysis.

Pre-response period. The resulting group-average power estimates of the pre-response period are visualised in Figure 4.2a. While all conditions appeared to show consistent ramping down in power approaching the time of response, there were clear differences between the characteristics of the response during trials of each condition. Beta-band power seemed to show the clearest distinction between conditions and less variability between subjects, however the same shape of response was also present in gamma-band power, and to a lesser extent in high-gamma band power with a slower, more gradual downward slope than the other two frequency bands.

The relative power of A trials appeared considerably closer to zero than V and AV trials within both beta- and gamma-band power, the former more-so than the latter, which could be related to the CPP electrodes used in this analysis being relatively close to visual processing areas compared to auditory, which would be positioned more laterally. This could also be related to the similarity of V and AV trial power if this signal were more reflective of visual inputs to evidence accumulation. Referring to gamma-band power, it also appeared that AV peak power resided somewhat in between V and A trials. Interestingly, AV trials appeared to have relatively lower high-gamma power across almost the whole period relative to the time of response, with V and A trials instead mirroring one another.

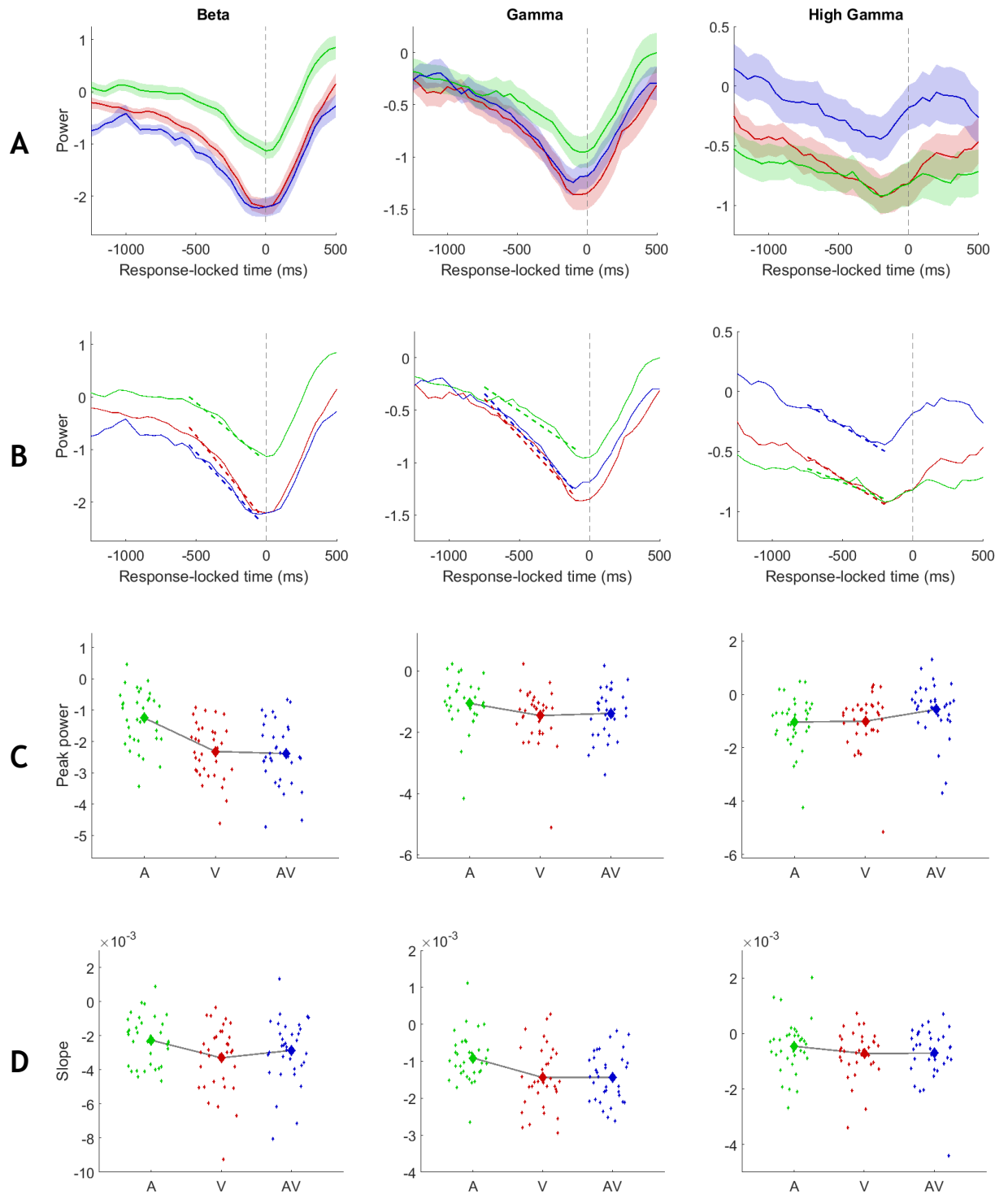


Figure 4.2 Spectral and statistical analysis results of period of interest leading up to the time of response ('pre-response' period). Colours indicate conditions as follows: red = visual trials (V), green = auditory trials (A), blue = audiovisual trials (AV). **A** Group mean across subject-specific peak frequencies per band (beta, gamma, high-gamma) of power estimates locked to the time of decision. Peak frequencies included per subject were selected using peak window timings specified in Table 4.1. Error bars are standard error of the mean. Dashed line indicates time of response. **B** The same group-averaged power estimates per condition, without error bars. Dashed coloured lines indicate the slope of the pre-response period of interest, using the mean slope (m) and peak time per condition calculated by averaging across subject-specific linear fits and peak times respectively. **C** Group mean of peak power per condition (larger points on grey line) with single-subject scatter. **D** Group mean across individual slopes (m), calculated in linear fits per condition and subject. Larger points are group means per condition. Smaller points scattered around these values are single-subject slopes per condition.

We examined the statistical differences in subject-specific peak power between conditions using three separate repeated-measures ANOVAs, one per frequency band, with each condition entered on a single level (Figure 4.2c). We also computed pairwise comparisons where a significant main effect was found, the results of which were adjusted using the Bonferroni adjustment for multiple comparisons, which provided a significance threshold of .017. First, for our comparison of beta-band power during this pre-response period of interest, Mauchly's test of sphericity was violated ($\chi^2(2) 17.04, p < .001$), and we accounted for this by correcting degrees of freedom using Greenhouse-Geisser estimates (adjusted $p = .708$). We found a significant main effect of sensory condition ($F(1.42, 46.72) = 30.92, p < .001$), with a large effect size as calculated using partial eta squared ($\eta^2 p = .48$). Pairwise comparisons showed significant differences between A trials and both V ($p < .001$) and AV trials ($p < .001$). We also found a significant main effect between sensory conditions in gamma-band peak power of medium effect size ($F(2, 66) = 3.07, p = .014, \eta^2 p = .12$; assumption of sphericity was met, $\chi^2(2) 1.08, p = .584$). Pairwise comparisons found no significant difference between V and A trials, according to the applied Bonferroni adjustment ($p = .036$), and no significant difference between A and AV trials ($p = .091$). Finally, there was a significant main effect of sensory condition regarding high-gamma-band power, also of medium effect size ($F(2, 66) = 4.57, p = .014, \eta^2 p = .12$; assumption of sphericity was met, $\chi^2(2) 2.28, p = .321$). However, this result was further decomposed into non-significant differences between A and AV trials ($p = .041$) as well as between V and AV trials ($p = .080$).

Next, we reviewed the differences in slope calculations between conditions (Figure 4.2b). Differences here were overall slightly more subtle, with some notable exceptions. Within beta-band power, A trials appeared to show the shallowest change in power over time, with V trial power ramping down at a much steeper slope, followed by AV trial power. However, AV trial power seemed to start from a lower power relative to V trials, which could explain some of this difference despite reaching roughly the same peak level. There were similar patterns in gamma-band power, with V and AV trial power ramping down at similar rates and both more steeply than A trial power. Finally, these

differences were mirrored in high-gamma-band power, however overall the rate of change appeared to be lower.

We examined the statistical difference between subject-specific ramps per condition and frequency band as we had for peak power, with three repeated-measures ANOVAs and Bonferroni-adjusted pairwise comparisons where a significant main effect was found (again, requiring $p < .017$). Firstly, our beta-band comparison violated the assumption of sphericity ($\chi^2(2) 10.37, p = .006$) therefore we adjusted our degrees of freedom using Greenhouse-Geisser estimates. We found a significant main effect of sensory condition on the slope of power, with a large effect size ($F(1.57, 51.70) = 6.32, p = .007, \eta^2p = .16$). Pairwise comparisons revealed that this difference was largely driven by a significant difference between V and A trials only ($P = .009$). Next, we found a significant main effect of sensory condition on gamma-band power with a large effect size ($F(2, 66) = 9.90, p < .001, \eta^2p = .23$; assumption of sphericity met, $\chi^2(2) 2.92, p = .232$). Pairwise comparisons revealed that A trial slopes were significantly different from both V trial ($P = .004$) and AV trial slopes ($P < .001$). Finally, we found no significant main effect of sensory condition on the slope of power within high-gamma-band frequencies ($F(2, 66) = 1.52, p = .227$; assumption of sphericity met, $\chi^2(2) 4.90, p = .086$).

Post-response period. Next, we moved on to look at differences between peak power in the period of rapid change at or shortly after the time of response. The resulting group-average, condition- and frequency-specific power estimates of this post-response period are visualised in Figure 4.3. The most notable difference when selecting peaks based on post-response power, rather than pre-response, was that gamma-band power seemed to show differences in power across time for AV trials compared to V and A trials (Figure 4.3a). This was in comparison to our first analysis where the difference in this frequency band seemed to be most prominent between A trials and V/AV trials. The overall profiles of beta-band and high-gamma-band power were largely the same, with A trial beta-band power substantially closer to zero throughout, and the same pronounced difference between AV trial high-gamma power and that of the other two unisensory conditions.

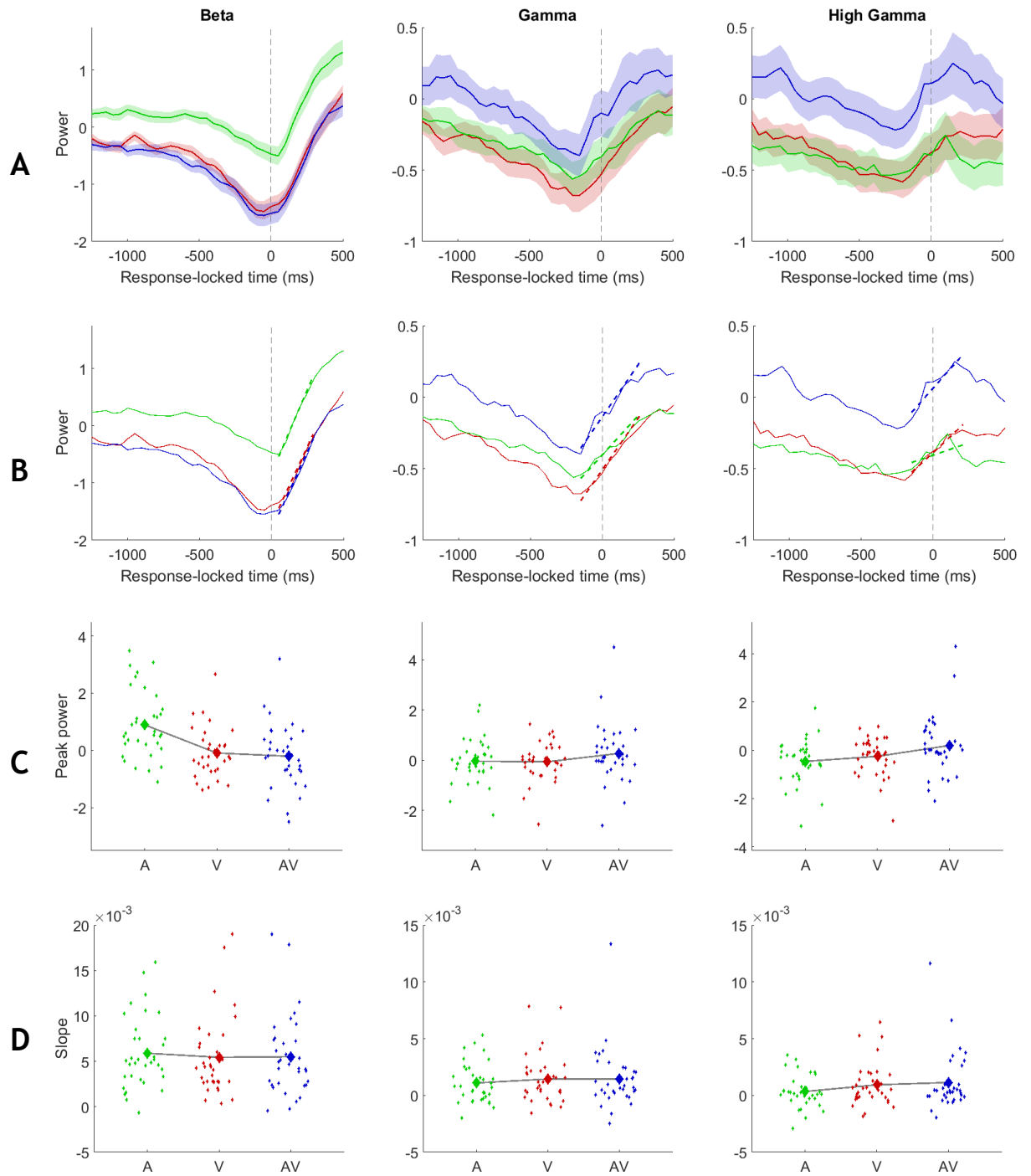


Figure 4.3 Spectral and statistical analysis results of period of interest at and shortly after the time of response ('post-response' period). Red = visual trials (V), green = auditory trials (A), blue = audiovisual trials (AV). A Group mean across subject-specific peak frequencies per band (beta, gamma, high-gamma) of power estimates locked to the time of decision. Peak frequencies included per subject were selected using peak window timings specified in Table 4.1. Error bars are standard error of the mean. Dashed line indicates time of response. B The same group-averaged power estimates per condition, without error bars. Dashed coloured lines indicate the slope of the pre-response period of interest, using the mean slope (m) and peak time per condition calculated by averaging across subject-specific linear fits and peak times respectively. C Group mean of peak power per condition (larger points on grey line) with single-subject scatter. D Group mean across individual slopes (m), calculated in linear fits per condition and subject. Larger points are group means per condition. Smaller points scattered around these values are single-subject slopes per condition.

We statistically compared, within each frequency band, the differences between condition-averaged peak power (extracted from timings specified in Table 4.1), using three separate repeated-measures ANOVAs. We also computed pairwise comparisons where a significant main effect was found, the results of which were adjusted using the Bonferroni adjustment for multiple comparisons (significance where $p < .017$). First, for our comparison of beta-band power, Mauchly's test of sphericity was violated ($\chi^2(2) 21.65$, $p < .001$), and we accounted for this by correcting degrees of freedom using Greenhouse-Geisser estimates (adjusted $p = .670$). We found a significant main effect of sensory condition ($F(1.34, 44.25) = 15.58$, $p < .001$), with a large effect size as calculated using partial eta squared ($\eta^2 p = .32$). Pairwise comparisons showed significant differences between A trials and both V ($p < .001$) and AV trials ($p = .001$). Despite the appearance of some differences regarding AV trials, we did not find a significant main effect of condition on gamma-band peak power, and therefore did not follow up with pairwise comparisons ($\chi^2(2) 7.06$, $p = .029$, Greenhouse-Geisser corrected $p = .835$; $F(1.67, 55.10) = 1.77$, $p = .178$). Finally, we found a significant main effect of sensory condition on high-gamma-band peak power; the assumption of sphericity was violated ($\chi^2(2) 11.73$, $p = .003$) therefore we corrected the degrees of freedom using Greenhouse-Geisser estimates ($p = .765$; $F(1.53, 50.50) = 5.62$, $p = .011$). Pairwise comparisons found a trend towards a difference in this case between A and AV trials ($p = .020$).

Next, we compared the linear fits estimating the rate of change in condition-specific power estimates during our window after the time of response (Table 4.1), within each frequency band (Figure 4.3b). Beta-band power appeared to increase at very similar rates across conditions, with the difference largely contained within power differences. However, A trial slopes appeared to be shallower than both V and AV trials in both gamma- and high-gamma-band power estimates. This roughly mirrored our findings from the pre-response portion of the analysis.

We examined the statistical difference between subject-specific ramps per condition and frequency band as we had for peak power, with three repeated-measures ANOVAs and Bonferroni-adjusted pairwise comparisons where a

significant main effect was found. Our beta-band comparison did not reveal a significant main effect of sensory condition on the rate of change in power (assumption of sphericity met, $\chi^2(2)$ 5.43, $p = .066$; $F(2, 66) = 0.36$, $p = .698$), nor did the comparison within gamma-band power (assumption of sphericity violated, $\chi^2(2)$ 5.43, $p = .005$; corrected with Greenhouse-Geisser estimates of degrees of freedom, $p = .782$; $F(1.56, 51.61) = .581$) or high-gamma-band power (assumption of sphericity violated, $\chi^2(2)$ 14.19, $p = .001$; corrected with Greenhouse-Geisser estimates of degrees of freedom, $p = .736$; $F(1.47, 48.60) = .155$).

To summarise our main findings: Shortly before the time of response we found that beta-band peak power estimates were significantly closer to zero during A trials compared to V or AV trials. Though gamma-band peak power estimates were affected by the type of sensory motion presented, pairwise comparisons could not reveal specific differences between conditions. Similarly, though there was a main effect of sensory modality on high-gamma-band peak power, pairwise comparisons did not reveal any significant differences between specific conditions. Additionally, the rate of change in power was significantly steeper during V trials than A trials within beta-band power estimates, and in V and AV trials than A trials within gamma-band power estimates. After the time of response, we found fewer significant differences overall; beta-band peak power was significantly higher during A trials than V and AV trials, though this seemed to be due to its overall smaller fluctuations away from zero over the course of the trial. Further, high-gamma-band peak power soon after the time of response was significantly affected by sensory condition. There were no significant differences in the rate of change in power after the time of response within any frequency band.

4.5 Discussion

In this chapter, we performed spectral analyses of EEG data recorded as participants completed a continuous audiovisual motion decision task. We computed time-frequency decompositions and investigated condition-specific differences between: a) the rate of change (ramping up or down) in power estimates approaching or near the time of response; and b) the subject-specific maximum or minimum power at these times. These were each computed within beta-band, gamma-band, and high-gamma-band power. In lower frequencies, pre-decision differences seemed to be present in that A trial power was closer to zero and changed more slowly than other sensory conditions, but in the highest frequencies AV trial power was considerably closer to zero than either unisensory condition which both reduced much further in power by the time of response. Our analysis of the period following the time of response revealed similar findings regarding the overall upward shift in A trial beta-band power, but it did not replicate the same difference between the change in power of these conditions. Selecting peaks based on the post-response period suggested that AV trial power was the outlying condition within gamma-band frequencies, however this difference was not significant. Nonetheless, as in the pre-response period, high-gamma power was significantly affected by sensory condition.

Overall, while we did find some interesting differences in power between conditions, they were not as consistent with some of the literature we described in the introduction as we might have hoped. We were most interested in changes relating to evidence accumulation, the setting of a decision boundary, or other aspects such as confidence which might have been visible as power modulation shortly after the time of response (Pleskac and Busemeyer, 2010; Fleming, Huijgen and Dolan, 2012; Gherman and Philiastides, 2015), in particular within the CPP-related sensors we defined in Chapter 3. This followed discovering increased rates of evidence accumulation during audiovisual decision making when performing an ERP-centred analysis. Specifically, we were investigating whether any of these changes would occur as a result of the presence of complementary audiovisual information, as opposed to unisensory visual or auditory information alone.

As we expected, the temporal evolution of power leading up to the time of response did seem to resemble a decision process, with power beginning to ramp up or down, peaking shortly before or at the time of response, then beginning to return to ‘normal’ after the response. This resembled that of Kelly and O’Connell (2013) and Polanía *et al.* (2014) who also saw this ramping of activity associated with a decision making process, and our results here also mirrored the ERP analysis of Chapter 3 (see 3.3.6) where we saw a similar evolution. Regarding changes to the rate of evidence accumulation in particular, we looked for whether the additional information present in AV trials (compared to V or A trials) caused an increase in the rate of change of our condition-specific power estimates. Our working assumption was that the general increase in power leading up to the time of response was indicative of sequential sampling of the sensory evidence within the stimuli presented (Forstmann, Ratcliff and Wagenmakers, 2016). Further, if more evidence were present in some trial types compared to others (i.e. AV trials compared to V or A trials), it is possible that the rate of evidence accumulation may have increased (Ratcliff *et al.*, 2016), and this difference could have been visible in the steepness of the slope of the activity related to evidence accumulation (as in Chapter 3), hence why we chose to quantify the slope of linear fits to rapidly-changing power and calculate the difference between condition-specific slopes. If AV trials led to an increased rate of evidence accumulation compared to trials presenting A or V motion alone, the slope of AV trial power may have been steeper, resembling a superadditive effect if demonstrably steeper than either of the unisensory conditions. This could suggest that the rate of evidence accumulation is enhanced by the additional sensory evidence present in AV trials. A major distinction in the approaches of these two chapters, however, is that our previous ERP-driven analysis was likely primarily driven by lower frequency activity, while here we focused on the “motor accumulation” activity reflected within higher frequencies (i.e., gamma- and high-gamma-band activity). This difference in perspective may not have captured activity in the same way using the CPP-related sensors as in the previous chapter.

However, we ultimately found that AV and V trial slopes were statistically indistinguishable within beta-band power changes measured here. This is mirrored in the average response times of these trials, which were also

comparable (Figure 3.2), further suggesting little difference in the temporal profiles of V and AV trial beta-band activity. The average slope of A trials was significantly shallower than both V and AV trials, mirroring our finding that response times were overall slower during A trials, which suggests a slower rate of evidence accumulation. This difference also mirrors our ERP findings in Chapter 3, suggesting that, in this task, the auditory condition was largely driving the effect.

As discussed in the previous chapter, some of the uncertainty in our measurement of evidence-accumulation-related neural activity may be due to the task design. The task deliberately makes the onset of coherent motion difficult to detect; no obvious cue to the switch from ITI to the target period is provided in any trial, and stimuli remain largely unchanged in appearance except for the proportion of dots moving together in any one direction. The intention was to reduce the effect of visually evoked potentials on the initial stages of evidence accumulation, however it is possible that the extra demand on participants to notice this change may have caused motion detection processing to be represented to a greater magnitude within the neural activity we recorded than expected. Direction discrimination is still an integral part of succeeding at the task, however this may have been relatively trivial to do compared to detecting that coherent motion was being presented and a response was required. The effects of this are most obvious in the high levels of direction discrimination accuracy once missed trials are removed, to the extent that we had to include all trials and instead look at modulation of the miss rate by coherence and sensory condition (and as Kelly & O'Connell did in their original analysis).

One takeaway, however, was that we did not see the same pattern of effect when calculating the rate of change of power in our 'post-response' period, which suggests that the differences we did discover were reliable in themselves and not the result of noise; if our analysis had been an unsuitable method to measure of the rate of evidence accumulation, there would not have been any distinguishable difference between the pre-response and post-response periods.

Another focus of our analysis was to investigate whether the complementary audiovisual information had an effect on peak power around the time of

response. A peak in neural activity shortly before a response is made may indicate the surpassing of a decision criterion and the enacting of said choice as required by the task at hand (e.g. saccade, button press). Other research has found modulations in peak power with increased sensory evidence or higher levels of confidence (Gherman and Philiastides, 2015). In the case of our experiment, we wanted to quantify the difference between condition-specific peak power to investigate whether complementary sensory evidence would impact on peak power (i.e. AV evidence versus A or V evidence alone).

Confidence is a key component of many real-world decisions, particularly those involving the accumulation of noisy sensory information such as when looking and listening for cars during foggy weather. In the face of environmental noise, confidence can be reflected in the amount of information required to be collected before a decision is made (i.e. the decision criterion). Multisensory sources of information are one way of increasing evidence availability, and could therefore increase decision confidence, particularly when that information is complementary in nature (Keane *et al.*, 2015). For example, if weather conditions are poor and it is objectively harder to tell if a car is approaching, your confidence in deciding whether to cross a road or not is likely to be higher if you can watch and listen for cars, compared to if you had to rely on sight or sound alone.

Our participants knew to anticipate multiple sensory sources of information during AV trials as we prompted them with this information before each AV block. Knowledge of this upcoming complementary information may have impacted on decision confidence, causing participants to approach AV trials with more confidence relative to V or A trials. In a study by Gherman and Philiastides (2015), levels of confidence seemed to modulate peak power either around the time of, or just after, response. Similarly, Philiastides *et al.* (2014) found scalp potentials near the time of response were modulated by the amount of sensory evidence available during the trial, and that inter-trial variability in this was predictive of behaviour. Due to the differing levels of confidence per condition that might have been induced by the varying amounts of sensory information present, we might therefore have expected to see increased peak power in AV trials compared to V or A trials. We were also inclined to think there would be a

difference as we knew that behavioural performance (i.e. miss rate) was enhanced during AV trials compared to V or A trials, meaning that there was a real difference in performance that might indicate increased confidence. In the case of our data, we compared condition-specific peak power both shortly before the time of response, specifically within our ‘pre-response’ window per frequency band, as well as within our ‘post-response’ window.

Comparing peak beta-band power revealed similar statistical differences to our ERP analysis of Chapter 3; V and AV trial power peaked at comparable levels, with A trial power peaking significantly closer to zero. However, in contrast to our ERP analysis, the peaks of V and AV power were considerably different shortly before the response, with AV trials peaking much closer to zero. In fact, when taking into account the similarity of the slope of power in this frequency band, there appears to be an overall modulation in high-gamma-band power, in that the pattern of change in high-gamma-power is very similar between conditions, but the ‘starting points’ for V and AV trials reflect the disparity in power seen throughout the trial. This same pattern of results was seen both in pre-response and post-response peak power (i.e. where different group-averaged power per condition was calculated based on the relative peak times per each analysis). Interestingly, gamma-band power almost shows the transition between these, with AV trial power appearing to peak closer to zero overall but not distinctly enough to be statistically significant.

It is possible that some of the differences between conditions seen here were reflected then in an overall offset in power, rather than strictly within the time up to and shortly after the response. To investigate this effect, a future analysis might include comparing peak-to-trough power amplitude, rather than comparing only the value at peak power itself across conditions. By doing so, we would account for overall differences in power that are more consistent across the trial, and highlight the change in power potentially caused by decision making activity shortly before the response.

One interesting feature of our power estimates is that we see an overall reduction in high-gamma power during AV trials, compared to A and V trial power. This could be a desynchronisation effect. Within all frequency bands analysed here, the desynchronisation of power seen may be indicative of a

contribution from premotor structures. Traditionally, these structures have been thought of as primarily executing the result of decision making, with evidence accumulation towards a decision criterion encapsulated in parietal and frontal regions. This information would then be projected forward to motor and premotor areas for instigation of the choice made. However, recent research has investigated the role of premotor structures within the decision process itself. Mice trained to complete a delayed match to sample olfactory task were significantly impaired during context-dependent decisions, but not those where context was not needed, following inactivation of downstream anterolateral motor cortex (Wu *et al.*, 2020), suggesting it encodes such contextual information for the task (as opposed to passively enacting a decision reached elsewhere). In a recent preprint (not peer-reviewed at the time of writing) involving macaques trained to discriminate between pattern orientations, inactivation of the superior colliculus was shown to alter the balance of evidence accumulation, as opposed to augmenting sensory or motor processing or by biasing evidence before accumulation (Jun *et al.*, 2020). Other researchers found that activity in the posterior medial frontal cortex (pmFC) during value-based decision making could not be purely explained by the upcoming motor response, but included further trial-by-trial decision dynamics, suggesting it encoded or embodied decision-related activity itself (Pisauro *et al.*, 2017). Importantly, this work also identified a centroparietal cluster linked to decision-related evidence accumulation, and trial-by-trial variability in signals from this cluster explained fMRI responses in pmFC, suggesting a link in decision-related activity between these regions. This may contribute to the desynchronisation seen in our results, should the pmFC be influencing the activity in the CCP measured in our study.

The decision evidence reflected in these areas seems to be more substantial than just ‘echoes’ of the activity taking place in areas that feed forward into them, rather they have a fundamental role in the formation of the decision itself. Theories of decision making processing hierarchies previously centred around a clear distinction between the sensory processing and evidence accumulation stages and the motor activity, decision-enacting stages of neural activity, however this more recent line of thinking suggests that aspects of the decision process are fed forward and embodied in motor areas while decision

making takes place. Recent works have begun to attempt to disentangle the relationship between these processes (Verdonck, Loossens and Philiastides, 2020), and importantly the influence of these regions has not been accounted for to a sufficient extent in prior sequential sampling works, as the newer evidence in favour of the embodiment hypothesis suggests (Ratcliff, 1978; Palmer, Huk and Shadlen, 2005).

Following this, future work might seek to further understand the role of motor effectors in decision processing, as characterised by sequential-sampling models, by manipulating the motor effector used and assessing how models differ in identifying regions linked to decision variables. For example, when participants make perceptual decisions using hand movements might be explained by the supplementary motor area (SMA), then asking them to perform the same task but using saccades might show models shift and link decision variables to the superior colliculus of the lateral intraparietal cortex (LIP, as in Jun *et al.*). This would build on Filimon *et al.* (2013), which disentangled perceptual decision making from motor processing by notifying participants of their response method (saccade or hand movement) only after stimuli had appeared, meaning participants would make a decision before being able to commit to a motor response, allowing researchers to look at these two stages of their overall response more independently. Repeating a similar experiment but using a neurally-informed drift diffusion model might help us to identify the specific roles of perceptual and motor processing in these regions, and their links to decision variables.

To conclude, despite some issues with the task design leading to an unexpected emphasis on the detection component of perceptual decision making, we were able to capture the temporal profile of evidence accumulation within beta, gamma, and high-gamma power estimates. The rate of change in power was generally greater in trials including visual motion compared to sound motion, however due to limitations identified in the previous chapter a more appropriately matched stimulus may need to be designed to make these more directly comparable, as visual and audiovisual power was often similar during decision processing in this task. Future work should also investigate the role of motor effectors in decision making more specifically, as the desynchronisation of

power seen here may be related to the currently popular embodiment hypothesis.

5 General Discussion

5.1 Overview

Perceptual decision making is an absolutely essential part of the way we navigate through life and happens so frequently, yet so seamlessly, despite often occurring in the presence of high levels of noise. Recognising the face of your own child in among a sea of other children leaving school, noticing the doorbell ring over the sound of playing music, or selecting the freshest-looking fruit in a supermarket; these are all processes that require the collection of sensory evidence in favour of one of many alternatives. However, the world is multisensory by nature and these decisions are often complicated further, but not necessarily hindered, by multiple types of sensory information relevant to the decision we are making. In this thesis, we have used the continued analogy of the decision to cross a road on a misty night as it exemplifies how mundane, but also how essential, the ability to combine such information is. Understanding the fundamental neural correlates of this process is therefore incredibly important, and even more so when we consider how the dysfunction of this process may impact on people's lives (Huang *et al.*, 2015; O'Callaghan *et al.*, 2017).

As discussed in the introductory chapter, there has been significant process towards understanding the neural underpinnings of unisensory perceptual decision making. Early progress in animal models of decision making implicated patterns of neural activity linked to evidence accumulation within specific regions of the brain (Gold and Shadlen, 2007). This formed the basis of sequential sampling models (Ratcliff, 1978), which have subsequently been used to characterise and decompose human behaviour using a wide array of perceptual decision tasks (Kelly and O'Connell, 2015). A well-developed field of neuroscientific research has now implicated areas such as LIP (Huk, Katz and Yates, 2017), the superior colliculus (Ratcliff, Cherian and Segraves, 2003), and the DLPFC (Heekeren *et al.*, 2006) in activity relating to evidence accumulation and decision formation. At the same time, researchers have worked to characterise the process of sensory integration, including feedforward and

feedback mechanisms between primary sensory cortices (Kayser and Logothetis, 2007). However, it is only in recent years that multisensory decision making has received as much research attention, therefore relatively little is known about the mechanisms underpinning it. In an ongoing discussion, the early integration (Schroeder and Foxe, 2005; Ghazanfar and Schroeder, 2006) and the late integration hypotheses (Bizley, Jones and Town, 2016) have argued for distinct temporal profiles within which sensory information is transformed into a resolved decision. In addition, there is still much to learn regarding how behavioural enhancements during multisensory decision making are reflected in neural activity.

To this end, the current thesis sought to understand further the neural correlates of multisensory decision making, specifically audiovisual decision making. We hoped to further characterise the temporal evolution of evidence accumulation towards multisensory decision formation, and see whether changes or enhancements in this activity can be linked with enhancements in behavioural performance.

5.2 Key findings

In Chapter 2, we provided evidence that post-sensory evidence accumulation of visual information was enhanced by the presence of complementary auditory evidence. Our approach capitalised on a well-established paradigm that had previously been used to investigate decision making behaviour; the face versus car task (Philiastides and Sajda, 2006b; Philiastides, Ratcliff and Sajda, 2006; Philiastides *et al.*, 2011) uses a linear discriminant analysis to identify activity specifically linked to decision making behaviour, and has been successfully employed to characterise the temporal evolution of decision processing, including perceptual learning (Diaz, Queirazza and Philiastides, 2017), and confidence (Gherman and Philiastides, 2015). By adapting this task to investigate audiovisual decision making, we could have confidence that the task would be properly capturing the relevant activity and form a solid basis for comparison between sensory conditions. This was done so successfully, and we were able to reveal that complementary sound stimuli enhanced late post-sensory decision

processing, but not early sensory encoding. This was in opposition to the early integration hypothesis which posits that information from more than one modality is combined earlier on in the decision, then a multisensory representation of evidence is accumulated towards a decision boundary. Instead, here we observed that the additional information changed late decision-processing, specifically the rate of evidence accumulation, as shown by a neurally-informed drift diffusion model. Interestingly, the scalp topography associated with this activity suggested that a centroparietal cluster of sensors contributed substantially to this process, in a similar fashion to neural representations of evidence accumulation demonstrated using other perceptual tasks and sensory modalities (O'Connell, Dockree and Kelly, 2012; Kelly and O'Connell, 2013; Philiastides, Heekeren and Sajda, 2014). Participants were more accurate in their decisions, and trial-by-trial behaviour could be predicted well by neural activity within our diffusion model. Overall, our results suggested that audiovisual enhancements of decision making activity occurred during post-sensory evidence accumulation leading up to the time of response.

Chapter 3 continued to investigate this pattern of neural activity by employing a modified version of a task that had been reported to capture the neural markers of evidence accumulation (Kelly and O'Connell, 2013) in a way that would prevent early visually-evoked potentials (VEPs) from obscuring observation and that might address concerns we had regarding how well audiovisual sensory information could be mapped onto each other in our previous task. Importantly, our analysis focused on a centroparietal positivity (CPP) component, documented by Kelly & O'Connell, but that also mirrored the topographical pattern of activity we observed in Chapter 2. Our results again revealed enhancements in the rate of evidence accumulation during audiovisual trials, compared to visual-alone (Drugowitsch *et al.*, 2014), with peak activity also reaching a somewhat higher level shortly before the time of response. We considered that this may be linked to increased levels of confidence during AV trials, however an attempt to retroactively study this was unsuccessful; a task that collected more direct reporting of decision confidence (Kiani, Corthell and Shadlen, 2014; Gherman and Philiastides, 2015, 2018) would have been a more appropriate measure of this. While the modulation of neural activity observed during AV trials generally followed the pattern expected, the strengths of these

effects were not necessarily as stark as we had considered. Importantly, a potential issue with the task discovered during our statistical analysis suggested that, while evidence accumulation was indeed being captured, a larger-than-expected representation of non-decision, detection-related activity may have made it harder to capture these effects than anticipated (see 5.3 below for further discussion). Nevertheless, our results overall supported the hypothesis that visual representations of sensory evidence were enhanced by auditory information during post-sensory evidence accumulation, which also supported the findings of Chapter 2.

In Chapter 4, we aimed to further investigate the neural underpinnings of this activity by decomposing the broadband signal collected in the electrophysiological data of Chapter 3 into specific frequencies of neural activity. We sought to understand whether additional information regarding the accumulation of multisensory evidence might be contained in oscillatory patterns of activity. Using the data collected in Chapter 3, we performed a spectral analysis and presented the temporal evolution of beta (Donner *et al.*, 2009; Siegel, Engel and Donner, 2011; O’Connell, Dockree and Kelly, 2012; Wyart *et al.*, 2012), gamma (Wang, 2002; Bollimunta and Ditterich, 2012; Polanía *et al.*, 2014), and high-gamma activity over the course of the trial. Unfortunately, this analysis failed to reveal the same robust differences in evidence accumulation (captured here as a gradual increase in power estimates leading up to response time) between AV and V trials. However, we reported what appeared to be an overall modulation in high-gamma power during AV trials, which may have reflected a contribution from premotor areas; the recent embodiment hypothesis posits that evidence accumulation signals can contain decision-preparatory activity independent of simple motor preparation (Jun *et al.*, 2020; Verdonck, Loossens and Philiastides, 2020; Wu *et al.*, 2020). We suggested that the desynchronisation we observed may have been linked to representations of decision-related activity in premotor areas during the task. We further suggested that future study might employ a perceptual task with delayed decision method instructions such as Jun *et al.*, in combination with a neurally-informed drift diffusion model, to differentiate areas associated with perceptual and motor processing.

5.3 Limitations and future directions

One immediate limitation of the current thesis is its reliance on data recorded using EEG alone. While it offers clear advantages in terms of the temporal resolution of neural activity it is possible to capture, its ability to capture the spatial distribution of activity is relatively limited. Here, our analysis either does not specify a region from which to extract activity (Chapter 2) or it does so based on the previously established CCP spatial component. This may have helped to limit the possibility of capturing irrelevant or confounding activity relating to our task, as we did not attempt to select a new region without good basis, however it does inherently limit the scope of what we were able to study in this case. Various other studies have implicated different neural regions in decision processing using methods such as fMRI, however our technique of choice meant we were unable to do so with so much clarity here. Future work may either use a different neuroimaging technique such as this to capture this information in a complementary way, or indeed aim to combine them using advanced techniques such as combined EEG/fMRI. From there specific regions implicated in evidence accumulation could be identified and the scope of the analysis focused in further. Some of our issues capturing a clear difference between sensory conditions may be resolved by doing so.

This change in approach may have been useful when investigating the role of premotor regions in oscillatory activity in particular; as we have already discussed, there is a growing suggestion that these regions may embody the decision signal in a way that is not simply motor preparation, and our results have made some suggestion in favour of this. We therefore suggest that further study could either aim to repeat a similar analysis but instead using a cluster from the premotor region, or indeed identify a task that would be better suited to capturing this effect.

In relation to this, a considerable limitation arose in the task used in Chapters 3 and 4. It appeared as though, while the task intended to allow an unimpeded view of the evidence accumulation process, this requirement to first detect coherent motion may have caused a disproportionate (compared to that

intended) representation of activity relating to detection earlier on in the trial. Our initial intention was to use accuracy, meaning the proportion of trials in which participants could successfully discriminate motion direction, as our behavioural measure of performance. However, we found that once we accounted for missed trials, participants very often performed at ceiling levels, suggesting that once they had detected that a coherent motion period was occurring, it was relatively trivial to then discriminate which direction that motion was travelling in. As we discussed in Chapter 3, this is noted somewhat by Kelly and O'Connell in their original paper, and in some subsequent works. One solution for this may be to modify the task in a relatively simple way; by prompting the participant to the start of a coherent motion period, a short while before that time begins, we may still be able to capture an unimpeded view of evidence accumulation during the trial, but participants are no longer required to perform detection to the same extent. In addition, we previously discussed a hope to investigate the effects of increased confidence during audiovisual trials. A further change to the task therefore might include asking participants to report decision confidence after all or some trials, and using this information to investigate whether increased confidence is indeed related to enhanced evidence accumulation in the same way we found audiovisual trials demonstrated here.

Finally, in a small but important note, my progress in compiling and reporting on the data collected for this thesis, as well as an up-to-date account of the literature surrounding it, was significantly disrupted by the COVID-19 pandemic. This has been a difficult period for almost everyone, however I feel that the specific circumstances around the thesis write-up portion of a PhD are less than complementary to additional major causes of isolation and stress (as well as others that had already occurred in my life). I was fortunate, however, to have completed data collection and the majority of my analysis before major changes began.

5.4 Conclusion

In summary, the current thesis has presented an account of the temporal evolution of evidence accumulation related activity during multisensory decision making, specifically audiovisual discrimination. It has shown that enhancements to visual representations of evidence during post sensory decision-making and evidence accumulation occur and are predictive of behavioural enhancements. Some of these effects were corroborated by subsequent study in Chapter 3, which showed a clear enhancement of evidence accumulation throughout the trial when complementary auditory evidence is provided. This was largely in support of previous literature demonstrating this effect, in opposition to those suggesting that the integration of audiovisual information during decision making occurs earlier on (i.e. the early integration hypothesis). An attempt to decompose the broadband signal of Chapter 3 into its constituent frequencies was less successful, however did highlight further opportunity for study in areas such as the role of premotor regions in the embodiment of evidence accumulation signals. Our results are a significant step towards understanding the temporal evolution of the neural correlates of multisensory perceptual decision making, and support the representation of this activity within centroparietal regions, however further study is needed to show the role of other areas such as premotor regions in the embodiment of evidence accumulation.

List of References

- Aller, M. and Noppeney, U. (2019) 'To integrate or not to integrate: Temporal dynamics of hierarchical Bayesian causal inference', *PLoS Biology*. Edited by C. Petkov. Public Library of Science, 17(4), p. e3000210. doi: 10.1371/journal.pbio.3000210.
- Angelaki, D. E., Gu, Y. and DeAngelis, G. C. (2009) 'Multisensory integration: psychophysics, neurophysiology, and computation.', *Current Opinion in Neurobiology*. Elsevier Current Trends, 19(4), pp. 452-8. doi: 10.1016/j.conb.2009.06.008.
- Audacity-Team (2016) 'Audacity'. Available at: <http://audacityteam.org/>.
- Baayen, R. H., Davidson, D. J. and Bates, D. M. (2008) 'Mixed-effects modeling with crossed random effects for subjects and items', *Journal of Memory and Language*. Academic Press, 59(4), pp. 390-412. doi: 10.1016/j.jml.2007.12.005.
- Barr, D. J. *et al.* (2013) 'Random effects structure for confirmatory hypothesis testing: Keep it maximal', *Journal of Memory and Language*. Academic Press, 68(3), pp. 255-278. doi: 10.1016/j.jml.2012.11.001.
- Bates, D. *et al.* (2015) 'Fitting linear mixed-effects models using lme4', *Journal of Statistical Software*. American Statistical Association, 67(1), pp. 1-48. doi: 10.18637/jss.v067.i01.
- Beauchamp, M. S. *et al.* (2004) 'Integration of auditory and visual information about objects in superior temporal sulcus', *Neuron*. Cell Press, 41(5), pp. 809-823. doi: 10.1016/S0896-6273(04)00070-4.
- Bernstein, L. E. *et al.* (2008) 'Spatiotemporal dynamics of audiovisual speech processing', *NeuroImage*. Academic Press, 39(1), pp. 423-435. doi: 10.1016/j.neuroimage.2007.08.035.
- Bernstein, L. E., Auer, E. T. and Takayanagi, S. (2004) 'Auditory speech detection in noise enhanced by lipreading', *Speech Communication*. North-

Holland, 44(1-4), pp. 5-18. Available at:

<https://www.sciencedirect.com/science/article/abs/pii/S0167639304001165>

(Accessed: 23 May 2019).

Bizley, J. K., Jones, G. P. and Town, S. M. (2016) 'Where are multisensory signals combined for perceptual decision-making?', *Current Opinion in Neurobiology*. Elsevier Ltd, 40, pp. 31-37. doi: 10.1016/j.conb.2016.06.003.

Blank, H. *et al.* (2013) 'Temporal characteristics of the influence of punishment on perceptual decision making in the human brain', *Journal of Neuroscience*. Society for Neuroscience, 33(9), pp. 3939-3952. doi: 10.1523/jneurosci.4151-12.2013.

Bogacz, R. *et al.* (2006) 'The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks', *Psychological Review*, 113(4), pp. 700-765. doi: 10.1037/0033-295X.113.4.700.

Bogacz, R. (2007) 'Optimal decision-making theories: linking neurobiology with behaviour', *Trends in Cognitive Sciences*. Elsevier Current Trends, 11(3), pp. 118-125. doi: 10.1016/j.tics.2006.12.006.

Bogacz, R. *et al.* (2010) 'The neural basis of the speed-accuracy tradeoff', *Trends in Neurosciences*. Elsevier Current Trends, 33(1), pp. 10-6. doi: 10.1016/j.tins.2009.09.002.

Bollimunta, A. and Ditterich, J. (2012) 'Local Computation of Decision-Relevant Net Sensory Evidence in Parietal Cortex', *Cerebral Cortex*. Oxford University Press, 22(4), pp. 903-917. doi: 10.1093/cercor/bhr165.

Brang, D. *et al.* (2013) 'Parietal connectivity mediates multisensory facilitation', *NeuroImage*. Elsevier Inc., 78, pp. 396-401. doi: 10.1016/j.neuroimage.2013.04.047.

Britten, K. H. *et al.* (1992) 'The analysis of visual motion: a comparison of neuronal and psychophysical performance.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*. Society for Neuroscience,

12(12), pp. 4745-65. doi: 10.1523/jneurosci.1655-06.2006.

Buzsáki, G. and Wang, X.-J. (2012) 'Mechanisms of Gamma Oscillations', *Annual Review of Neuroscience*. Annual Reviews , 35(1), pp. 203-225. doi: 10.1146/annurev-neuro-062111-150444.

Cao, Y. *et al.* (2019) 'Causal Inference in the Multisensory Brain', *Neuron*. Cell Press, 102(5), pp. 1076-1087.e8. doi: 10.1016/j.neuron.2019.03.043.

Carland, M. A. *et al.* (2016) 'Evidence against perfect integration of sensory information during perceptual decision making.', *Journal of neurophysiology*. American Physiological Society, 115(2), pp. 915-30. doi: 10.1152/jn.00264.2015.

Cavanagh, J. F. *et al.* (2011) 'Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold', *Nature Neuroscience*. Nature Publishing Group, 14(11), pp. 1462-1467. doi: 10.1038/nn.2925.

Cavanagh, J. F. *et al.* (2014) 'Eye tracking and pupillometry are indicators of dissociable latent decision processes', *Journal of Experimental Psychology: General*. American Psychological Association Inc., 143(4), pp. 1476-1488. doi: 10.1037/a0035813.

Chandrasekaran, C. (2017) 'Computational principles and models of multisensory integration.', *Current Opinion in Neurobiology*. NIH Public Access, 43, pp. 25-34. doi: 10.1016/j.conb.2016.11.002.

Chandrasekaran, C., Lemus, L. and Ghazanfar, A. A. (2013) 'Dynamic faces speed up the onset of auditory cortical spiking responses during vocal detection', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 110(48), pp. E4668-E4677. doi: 10.1073/pnas.1312518110.

Chen, Y.-C. *et al.* (2011) 'Synchronous sounds enhance visual sensitivity without reducing target uncertainty.', *Seeing and perceiving*. Brill, 24(6), pp. 623-38. doi: 10.1163/187847611X603765.

Clark, J. and Yuille, A. (1990) *Data fusion for sensory information processing*

systems. Boston: Kluwer Academic.

Cohen, M. X. (2014) *Analyzing Neural Time Series Data*. MIT Press 238 Main St., Suite 500, Cambridge, MA. Available at: <https://mitpress.mit.edu/books/analyzing-neural-time-series-data> (Accessed: 15 March 2021).

Colonus, H. and Diederich, A. (2018) 'Formal models and quantitative measures of multisensory integration: A selective overview', *European Journal of Neuroscience*. John Wiley & Sons, Ltd (10.1111), 51(5), pp. 1161-1178. doi: 10.1111/ejn.13813.

Cooper, N. R. *et al.* (2003) 'Paradox lost? Exploring the role of alpha oscillations during externally vs. internally directed attention and the implications for idling and inhibition hypotheses', *International Journal of Psychophysiology*. Elsevier, 47(1), pp. 65-74. doi: 10.1016/S0167-8760(02)00107-1.

Dakin, S. . *et al.* (2002) 'What causes non-monotonic tuning of fMRI response to noisy images?', *Current Biology*, 12(14), pp. R476-R477. doi: 10.1016/S0960-9822(02)00960-0.

Delis, I. *et al.* (2016) 'Space-by-time decomposition for single-trial decoding of M/EEG activity', *NeuroImage*. Elsevier, 133, pp. 504-515. Available at: <https://www.sciencedirect.com/science/article/pii/S1053811916002482> (Accessed: 2 April 2019).

Delis, I. *et al.* (2018) 'Correlation of neural activity with behavioral kinematics reveals distinct sensory encoding and evidence accumulation processes during active tactile sensing', *NeuroImage*. Academic Press, 175, pp. 12-21. doi: 10.1016/j.neuroimage.2018.03.035.

Diaz, J. A., Queirazza, F. and Philiastides, M. G. (2017) 'Perceptual learning alters post-sensory processing in human decision-making', *Nature Human Behaviour*. Nature Publishing Group, 1(2), p. 35. doi: 10.1038/s41562-016-0035.

Ding, L. and Gold, J. I. (2010) 'Caudate encodes multiple computations for

perceptual decisions.’, *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(47), pp. 15747-59. doi: 10.1523/JNEUROSCI.2894-10.2010.

Ding, L. and Gold, J. I. (2012) ‘Neural Correlates of Perceptual Decision Making before, during, and after Decision Commitment in Monkey Frontal Eye Field’, *Cerebral Cortex*. Oxford Academic, 22(5), pp. 1052-1067. doi: 10.1093/cercor/bhr178.

Ding, L. and Gold, J. I. (2013) ‘The Basal Ganglia’s Contributions to Perceptual Decision Making’, *Neuron*, 79(4), pp. 640-649. doi: 10.1016/j.neuron.2013.07.042.

Dmochowski, J. P. and Norcia, A. M. (2015) ‘Cortical components of reaction-time during perceptual decisions in humans’, *PLOS ONE*. Edited by L. Cattaneo. Public Library of Science, 10(11), p. e0143339. doi: 10.1371/journal.pone.0143339.

Donner, T. H. *et al.* (2009) ‘Buildup of Choice-Predictive Activity in Human Motor Cortex during Perceptual Decision Making’, *Current Biology*. Cell Press, 19(18), pp. 1581-1585. doi: 10.1016/j.cub.2009.07.066.

Drugowitsch, J. *et al.* (2012) ‘The cost of accumulating evidence in perceptual decision making’, *Journal of Neuroscience*. Society for Neuroscience, 32(11), pp. 3612-3628. doi: 10.1523/JNEUROSCI.4010-11.2012.

Drugowitsch, J. *et al.* (2014) ‘Optimal multisensory decision-making in a reaction-time task’, *eLife*. eLife Sciences Publications Ltd, 3(3), p. 3005. doi: 10.7554/eLife.03005.

Drugowitsch, J. *et al.* (2015) ‘Tuning the speed-accuracy trade-off to maximize reward rate in multisensory decision-making’, *eLife*. eLife Sciences Publications Ltd, 4, p. e06678. doi: 10.7554/eLife.06678.

Duda, R., Hart, P. and Stork, D. (2001) *Pattern classification*, John Wiley & Sons.

- Duncan-Johnson, C. C. and Donchin, E. (1982) 'The P300 component of the event-related brain potential as an index of information processing', *Biological Psychology*, 14(1-2), pp. 1-52. doi: 10.1016/0301-0511(82)90016-3.
- Eckert, M. A. *et al.* (2008) 'A cross-modal system linking primary auditory and visual cortices: Evidence from intrinsic fMRI connectivity analysis', *Human Brain Mapping*. John Wiley & Sons, Ltd, 29(7), pp. 848-857. doi: 10.1002/hbm.20560.
- Esposito, F., Mulert, C. and Goebel, R. (2009) 'Combined distributed source and single-trial EEG-fMRI modeling: Application to effortful decision making processes', *NeuroImage*. Academic Press, 47(1), pp. 112-121. doi: 10.1016/j.neuroimage.2009.03.074.
- Falchier, A. *et al.* (2010) 'Projection from visual areas V2 and prostriata to caudal auditory cortex in the monkey', *Cerebral Cortex*. Narnia, 20(7), pp. 1529-1538. doi: 10.1093/cercor/bhp213.
- Filimon, F. *et al.* (2013) 'How Embodied Is Perceptual Decision Making? Evidence for Separate Processing of Perceptual and Motor Decisions', *Journal of Neuroscience*. Society for Neuroscience, 33(5), pp. 2121-2136. doi: 10.1523/jneurosci.2334-12.2013.
- FitzGerald, T. H. B. *et al.* (2015) 'Precision and neuronal dynamics in the human posterior parietal cortex during evidence accumulation', *NeuroImage*. Academic Press Inc., 107, pp. 219-228. doi: 10.1016/j.neuroimage.2014.12.015.
- Fleming, S. M., Huijgen, J. and Dolan, R. J. (2012) 'Prefrontal contributions to metacognition in perceptual decision making', *Journal of Neuroscience*. Society for Neuroscience, 32(18), pp. 6117-6125. doi: 10.1523/JNEUROSCI.6489-11.2012.
- Forstmann, B. U., Ratcliff, R. and Wagenmakers, E.-J. (2016) 'Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions', *Annual Review of Psychology*. Annual Reviews Inc., 67(1), pp. 641-666. doi: 10.1146/annurev-psych-122414-033645.
- Frank, M. J. *et al.* (2015) 'fMRI and EEG Predictors of Dynamic Decision

Parameters during Human Reinforcement Learning’, *Journal of Neuroscience*. Society for Neuroscience, 35(2), pp. 485-494. doi: 10.1523/JNEUROSCI.2036-14.2015.

Franzen, L. *et al.* (2020) ‘Auditory information enhances post-sensory visual evidence during rapid multisensory decision-making’, *Nature Communications*. Nature Research, 11(1), pp. 1-14. doi: 10.1038/s41467-020-19306-7.

Ghazanfar, A. A. and Schroeder, C. E. (2006) ‘Is neocortex essentially multisensory?’, *Trends in Cognitive Sciences*. Elsevier Current Trends, 10(6), pp. 278-285. doi: 10.1016/j.tics.2006.04.008.

Gherman, A. S. (2017) *Spatiotemporal neural correlates of confidence in perceptual decision making*, Thesis. University of Glasgow. Available at: <https://eleanor.lib.gla.ac.uk/record=b3287116> (Accessed: 15 March 2021).

Gherman, S. and Philiastides, M. G. (2015) ‘Neural representations of confidence emerge from the process of decision formation during perceptual choices’, *NeuroImage*. Elsevier Inc., 106, pp. 134-143. doi: 10.1016/j.neuroimage.2014.11.036.

Gherman, S. and Philiastides, M. G. (2018) ‘Human VMPFC encodes early signatures of confidence in perceptual decisions’, *eLife*. eLife Sciences Publications Ltd, 7, pp. 1-28. doi: 10.7554/eLife.38293.

Giard, M. H. and Peronnet, F. (1999) ‘Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study.’, *Journal of Cognitive Neuroscience*. MIT Press 238 Main St., Suite 500, Cambridge, MA 02142-1046 USA journals-info@mit.edu, 11(5), pp. 473-490. doi: 10.1162/089892999563544.

Gleiss, S. and Kayser, C. (2012) ‘Audio-Visual Detection Benefits in the Rat’, *PLoS ONE*. Edited by T. Boraud. Public Library of Science, 7(9), p. e45677. doi: 10.1371/journal.pone.0045677.

Gleiss, S. and Kayser, C. (2013) ‘Oscillatory mechanisms underlying the

enhancement of visual motion perception by multisensory congruency', *Neuropsychologia*. Elsevier, 53, pp. 84-93. doi: 10.1016/j.neuropsychologia.2013.11.005.

Gleiss, S. and Kayser, C. (2014) 'Acoustic noise improves visual perception and modulates occipital oscillatory States.', *Journal of Cognitive Neuroscience*, 26, pp. 699-711. doi: 10.1162/jocn.

Gold, Joshua I and Shadlen, M. N. (2007) 'The neural basis of decision making.', *Annual review of neuroscience*. Annual Reviews, 30(30), pp. 535-74. doi: 10.1146/annurev.neuro.29.051605.113038.

Green, D. M. and Von Gierke, S. M. (1984) 'Visual and auditory choice reaction times', *Acta Psychologica*. North-Holland, 55(3), pp. 231-247. doi: 10.1016/0001-6918(84)90043-X.

Green, D. M. and Swets, J. A. (1966) *Signal detection theory and psychophysics*. Wiley New York.

Guggenmos, M., Sterzer, P. and Cichy, R. M. (2018) 'Multivariate pattern analysis for MEG: A comparison of dissimilarity measures', *NeuroImage*. Academic Press, 173, pp. 434-447. doi: 10.1016/j.neuroimage.2018.02.044.

Haegens, S. *et al.* (2011) 'Beta oscillations in the monkey sensorimotor network reflect somatosensory decision making', *Proceedings of the National Academy of Sciences of the United States of America*. PNAS, 108(26), pp. 10708-10713. doi: 10.1073/pnas.1107297108.

Hauser, C. K. and Salinas, E. (2014) 'Perceptual Decision Making', in *Encyclopedia of Computational Neuroscience*. New York, NY: Springer New York, pp. 1-21. doi: 10.1007/978-1-4614-7320-6_317-1.

Heekeren, H. R. *et al.* (2004) 'A general mechanism for perceptual decision-making in the human brain', *Nature*. Nature Publishing Group, 431(7010), pp. 859-862. doi: 10.1038/nature02966.

- Heekeren, H. R. *et al.* (2006) 'Involvement of human left dorsolateral prefrontal cortex in perceptual decision making is independent of response modality', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 103(26), pp. 10023-10028. doi: 10.1073/pnas.0603949103.
- Herring, J. D. *et al.* (2015) 'Attention modulates TMS-locked alpha oscillations in the visual cortex', *Journal of Neuroscience*. Society for Neuroscience, 35(43), pp. 14435-14447. doi: 10.1523/JNEUROSCI.1833-15.2015.
- Herz, D. M. *et al.* (2016) 'Neural correlates of decision thresholds in the human subthalamic nucleus', *Current Biology*. Elsevier, 26(7), pp. 916-920. doi: 10.1016/j.cub.2016.01.051.
- Hirokawa, J. *et al.* (2011) 'Multisensory information facilitates reaction speed by enlarging activity difference between superior colliculus hemispheres in rats', *PLoS ONE*. Edited by H. D. Mansvelder. Public Library of Science, 6(9), p. e25283. doi: 10.1371/journal.pone.0025283.
- Ho, T. C., Brown, S. and Serences, J. T. (2009) 'Domain General Mechanisms of Perceptual Decision Making in Human Cortex', *Journal of Neuroscience*, 29(27), pp. 8675-8687. doi: 10.1523/JNEUROSCI.5984-08.2009.
- Huang, Y. T. *et al.* (2015) 'Different effects of dopaminergic medication on perceptual decision-making in Parkinson's disease as a function of task difficulty and speed-accuracy instructions', *Neuropsychologia*. Elsevier Ltd, 75, pp. 577-587. doi: 10.1016/j.neuropsychologia.2015.07.012.
- Huk, A. C., Katz, L. N. and Yates, J. L. (2017) 'The Role of the Lateral Intraparietal Area in (the Study of) Decision Making', *Annual Review of Neuroscience*. Annual Reviews Inc., 40, pp. 349-372. doi: 10.1146/annurev-neuro-072116-031508.
- Iurilli, G. *et al.* (2012) 'Sound-Driven Synaptic Inhibition in Primary Visual Cortex', *Neuron*. Cell Press, 73(4), pp. 814-828. doi: 10.1016/j.neuron.2011.12.026.

Jun, E. J. *et al.* (2020) 'Embodied cognition and decision-making in the primate superior colliculus', *bioRxiv*. bioRxiv, p. 2020.08.14.251777. doi: 10.1101/2020.08.14.251777.

Kayser, C. *et al.* (2007) 'Functional imaging reveals visual modulation of specific fields in auditory cortex.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(8), pp. 1824-35. doi: 10.1523/JNEUROSCI.4737-06.2007.

Kayser, C. and Logothetis, N. K. (2007) 'Do early sensory cortices integrate cross-modal information?', *Brain Structure and Function*. Springer-Verlag, 212(2), pp. 121-132. doi: 10.1007/s00429-007-0154-0.

Kayser, J. and Tenke, C. E. (2006) 'Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: I. Evaluation with auditory oddball tasks', *Clinical Neurophysiology*. Elsevier, 117(2), pp. 348-368. doi: 10.1016/j.clinph.2005.08.034.

Kayser, S. J., Philiastides, M. G. and Kayser, C. (2017) 'Sounds facilitate visual motion discrimination via the enhancement of late occipital visual representations', *NeuroImage*. Neuroimage, 148, pp. 31-41. doi: 10.1016/j.neuroimage.2017.01.010.

Keane, B. *et al.* (2015) 'Perceptual confidence demonstrates trial-by-trial insight into the precision of audio-visual timing encoding', *Consciousness and Cognition*. Academic Press Inc., 38, pp. 107-117. doi: 10.1016/j.concog.2015.10.010.

Kelly, S. P. and O'Connell, R. G. (2013) 'Internal and External Influences on the Rate of Sensory Evidence Accumulation in the Human Brain', *Journal of Neuroscience*, 33(50), pp. 19434-19441. doi: 10.1523/JNEUROSCI.3355-13.2013.

Kelly, S. P. and O'Connell, R. G. (2015) 'The neural processes underlying perceptual decision making in humans: Recent progress and future directions', *Journal of Physiology Paris*. Elsevier, 109(1-3), pp. 27-37. doi: 10.1016/j.jphysparis.2014.08.003.

- Van Kempen, J. *et al.* (2019) 'Behavioural and neural signatures of perceptual decision-making are modulated by pupil-linked arousal', *eLife*. eLife Sciences Publications Ltd, 8. doi: 10.7554/eLife.42541.
- Kiani, R., Corthell, L. and Shadlen, M. N. (2014) 'Choice certainty is informed by both evidence and decision time', *Neuron*, 84(6), pp. 1329-1342. doi: 10.1016/j.neuron.2014.12.015.
- Kiani, R. and Shadlen, M. N. (2009) 'Representation of confidence associated with a decision by neurons in the parietal cortex.', *Science (New York, N.Y.)*, 324(5928), pp. 759-64. doi: 10.1126/science.1169405.
- Klinge, C. *et al.* (2010) 'Corticocortical Connections Mediate Primary Visual Cortex Responses to Auditory Stimulation in the Blind', *Journal of Neuroscience*. Soc Neuroscience, 30(38), pp. 12798-12805. doi: 10.1523/jneurosci.2384-10.2010.
- Körding, K. P. *et al.* (2007) 'Causal inference in multisensory perception', *PLoS ONE*. Edited by O. Sporns. Public Library of Science, 2(9), p. e943. doi: 10.1371/journal.pone.0000943.
- Krajbich, I. and Rangel, A. (2011) 'Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 108(33), pp. 13852-13857. doi: 10.1073/pnas.1101328108.
- Kruschke, J. K. (2010a) 'Bayesian data analysis', *Wiley Interdisciplinary Reviews: Cognitive Science*. John Wiley & Sons, Ltd, 1(5), pp. 658-676. doi: 10.1002/wcs.72.
- Kruschke, J. K. (2010b) 'What to believe: Bayesian methods for data analysis', *Trends in Cognitive Sciences*. Elsevier Current Trends, 14(7), pp. 293-300. doi: 10.1016/j.tics.2010.05.001.
- Kulahci, I. G., Dornhaus, A. and Papaj, D. R. (2008) 'Multimodal signals enhance decision making in foraging bumble-bees.', *Proceedings. Biological sciences* /

The Royal Society, 275(April), pp. 797-802. doi: 10.1098/rspb.2007.1176.

Lakatos, P. *et al.* (2007) 'Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex', *Neuron*. Cell Press, 53(2), pp. 279-292. doi: 10.1016/j.neuron.2006.12.011.

Lange, J., Christian, N. and Schnitzler, A. (2013) 'Audio-visual congruency alters power and coherence of oscillatory activity within and between cortical areas', *NeuroImage*. Elsevier Inc., 79, pp. 111-120. doi: 10.1016/j.neuroimage.2013.04.064.

von Lautz, A., Herding, J. and Blankenburg, F. (2019) 'Neuronal signatures of a random-dot motion comparison task', *NeuroImage*. Academic Press, 193, pp. 57-66. doi: 10.1016/j.neuroimage.2019.02.071.

Leo, F. *et al.* (2011) 'Looming sounds enhance orientation sensitivity for visual stimuli on the same side as such sounds', *Experimental Brain Research*. Springer-Verlag, 213(2-3), pp. 193-201. doi: 10.1007/s00221-011-2742-8.

Lippert, M., Logothetis, N. K. and Kayser, C. (2007) 'Improvement of visual contrast detection by a simultaneous sound', *Brain Research*. Elsevier, 1173(1), pp. 102-109. doi: 10.1016/j.brainres.2007.07.050.

Maris, E. and Oostenveld, R. (2007) 'Nonparametric statistical testing of EEG- and MEG-data', *Journal of Neuroscience Methods*. Elsevier, 164(1), pp. 177-190. doi: 10.1016/j.jneumeth.2007.03.024.

McGovern, D. P. *et al.* (2018) 'Reconciling age-related changes in behavioural and neural indices of human perceptual decision-making', *Nature Human Behaviour*. Nature Publishing Group, 2(12), pp. 955-966. doi: 10.1038/s41562-018-0465-6.

Mercier, M. R. *et al.* (2013) 'Auditory-driven phase reset in visual cortex: Human electrocorticography reveals mechanisms of early multisensory integration', *NeuroImage*. Elsevier Inc., 79, pp. 19-29. doi: 10.1016/j.neuroimage.2013.04.060.

Mercier, M. R. and Cappe, C. (2020) 'The interplay between multisensory integration and perceptual decision making', *NeuroImage*. Academic Press Inc., 222, p. 116970. doi: 10.1016/j.neuroimage.2020.116970.

Mitra, P. (2007) *Observed brain dynamics*. Oxford University Press.

Mulder, M. J. *et al.* (2013) 'The speed and accuracy of perceptual decisions in a random-tone pitch task', *Attention, Perception, and Psychophysics*, 75(5), pp. 1048-1058. doi: 10.3758/s13414-013-0447-8.

Naci, L. *et al.* (2012) 'Are the senses enough for sense? Early high-level feedback shapes our comprehension of multisensory objects.', *Frontiers in integrative neuroscience*, 6, p. 82. doi: 10.3389/fnint.2012.00082.

Newman, D. P. *et al.* (2017) 'Visuospatial asymmetries arise from differences in the onset time of perceptual evidence accumulation', *Journal of Neuroscience*. Society for Neuroscience, 37(12), pp. 3378-3385. doi: 10.1523/JNEUROSCI.3512-16.2017.

Newsome, W. T., Britten, K. H. and Movshon, J. A. (1989) 'Neuronal correlates of a perceptual decision', *Nature*. Nature Publishing Group, 341(6237), pp. 52-54. doi: 10.1038/341052a0.

Newsome, W. T. and Pare, E. B. (1988) 'A selective impairment of motion perception following lesions of the middle temporal visual area (MT)', *Journal of Neuroscience*. Society for Neuroscience, 8(6), pp. 2201-2211. doi: 10.1523/jneurosci.08-06-02201.1988.

Noppeney, U., Ostwald, D. and Werner, S. (2010) 'Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex', *Journal of Neuroscience*. Society for Neuroscience, 30(21), pp. 7434-7446. doi: 10.1523/jneurosci.0455-10.2010.

Nunez, M. D., Srinivasan, R. and Vandekerckhove, J. (2015) 'Individual differences in attention influence perceptual decision making', *Frontiers in Psychology*. Frontiers Research Foundation, 6(FEB), p. 18. doi:

10.3389/fpsyg.2015.00018.

Nunez, M. D., Vandekerckhove, J. and Srinivasan, R. (2017) 'How attention influences perceptual decision making: Single-trial EEG correlates of drift-diffusion model parameters', *Journal of Mathematical Psychology*. Academic Press, 76, pp. 117-130. doi: 10.1016/j.jmp.2016.03.003.

O'Callaghan, C. *et al.* (2017) 'Visual Hallucinations Are Characterized by Impaired Sensory Evidence Accumulation: Insights From Hierarchical Drift Diffusion Modeling in Parkinson's Disease', *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. Elsevier Inc, 2(8), pp. 680-688. doi: 10.1016/j.bpsc.2017.04.007.

O'Connell, R. G., Dockree, P. M. and Kelly, S. P. (2012) 'A supramodal accumulation-to-bound signal that determines perceptual decisions in humans', *Nature Neuroscience*. Nature Publishing Group, 15(12), pp. 1729-1735. doi: 10.1038/nn.3248.

Oostenveld, R. *et al.* (2011) 'FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data', *Computational intelligence and neuroscience*. Hindawi, 2011.

Otto, T. U. and Mamassian, P. (2012) 'Noise and correlations in parallel perceptual decision making', *Current Biology*. Cell Press, 22(15), pp. 1391-1396. doi: 10.1016/j.cub.2012.05.031.

Palmer, J., Huk, a. C. and Shadlen, M. N. (2005) 'The effect of stimulus strength on the speed and accuracy of a perceptual decision', *Journal of Vision*, 5(5), pp. 1-1. doi: 10.1167/5.5.1.

Parra, L. *et al.* (2002) 'Linear spatial integration for single-trial detection in encephalography', *NeuroImage*, 17(1), pp. 223-230. doi: 10.1006/nimg.2002.1212.

Parra, L. C. *et al.* (2005) 'Recipes for the linear analysis of EEG', *NeuroImage*. Academic Press, 28(2), pp. 326-341. doi: 10.1016/j.neuroimage.2005.05.032.

Pedersen, M. L., Frank, M. J. and Biele, G. (2017) 'The drift diffusion model as the choice rule in reinforcement learning', *Psychonomic Bulletin and Review*. Springer, 24(4), pp. 1234-1251. doi: 10.3758/s13423-016-1199-y.

Peirce, J. *et al.* (2019) 'PsychoPy2: Experiments in behavior made easy', *Behavior Research Methods*. Springer US, 51(1), pp. 195-203. doi: 10.3758/s13428-018-01193-y.

Peirce, J. W. (2008) 'Generating stimuli for neuroscience using PsychoPy', *Frontiers in Neuroinformatics*. Frontiers Media S.A., 2, p. 10. doi: 10.3389/neuro.11.010.2008.

Pernet, C. R., Wilcox, R. and Rousselet, G. A. (2013) 'Robust correlation analyses: False positive and power validation using a new open source matlab toolbox', *Frontiers in Psychology*. Frontiers, 3(JAN), p. 606. doi: 10.3389/fpsyg.2012.00606.

Perrodin, C. *et al.* (2015) 'Natural asynchronies in audiovisual communication signals regulate neuronal multisensory interactions in voice-sensitive cortex.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 112(1), pp. 273-8. doi: 10.1073/pnas.1412817112.

Petro, L. S., Paton, A. T. and Muckli, L. (2017) 'Contextual modulation of primary visual cortex by auditory signals', *Philosophical Transactions of the Royal Society B: Biological Sciences*. The Royal Society, 372(1714), p. 20160104. doi: 10.1098/rstb.2016.0104.

Philiastides, M. G. *et al.* (2011) 'Causal role of dorsolateral prefrontal cortex in human perceptual decision making', *Current Biology*. Cell Press, 21(11), pp. 980-983. doi: 10.1016/j.cub.2011.04.034.

Philiastides, Marios G., Heekeren, H. R. and Sajda, P. (2014) 'Human scalp potentials reflect a mixture of decision-related signals during perceptual choices', *Journal of Neuroscience*. Society for Neuroscience, 34(50), pp. 16877-16889. doi: 10.1523/JNEUROSCI.3012-14.2014.

Philiastides, M. G., Ratcliff, R. and Sajda, P. (2006) 'Neural representation of task difficulty and decision making during perceptual categorization: A timing diagram', *Journal of Neuroscience*. Society for Neuroscience, 26(35), pp. 8965-8975. doi: 10.1523/JNEUROSCI.1655-06.2006.

Philiastides, M. G. and Sajda, P. (2006a) 'Causal influences in the human brain during face discrimination: A short-window directed transfer function approach', *IEEE Transactions on Biomedical Engineering*, 53(12), pp. 2602-2605. doi: 10.1109/TBME.2006.885122.

Philiastides, M. G. and Sajda, P. (2006b) 'Temporal characterization of the neural correlates of perceptual decision making in the human brain', *Cerebral Cortex*. Oxford University Press, 16(4), pp. 509-518. doi: 10.1093/cercor/bhi130.

Philiastides, M. G. and Sajda, P. (2007) 'EEG-informed fMRI reveals spatiotemporal characteristics of perceptual decision making', *Journal of Neuroscience*. Society for Neuroscience, 27(48), pp. 13082-13091. doi: 10.1523/jneurosci.3540-07.2007.

Pisauro, M. A. *et al.* (2017) 'Neural correlates of evidence accumulation during value-based decisions revealed via simultaneous EEG-fMRI', *Nature Communications*. Nature Research, 8(1), pp. 1-9. doi: 10.1038/ncomms15808.

Pleskac, T. J. and Busemeyer, J. R. (2010) 'Two-stage dynamic signal detection: a theory of choice, decision time, and confidence.', *Psychological review*. American Psychological Association, 117(3), p. 864.

Ploran, E. J. *et al.* (2007) 'Evidence accumulation and the moment of recognition: Dissociating perceptual recognition processes using fMRI', *Journal of Neuroscience*. Soc Neuroscience, 27(44), pp. 11912-11924. doi: 10.1523/JNEUROSCI.3522-07.2007.

Plummer, M. (2003) 'JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling', in *Proceedings of the 3rd international workshop on distributed statistical computing*. Vienna, Austria.

- Polanía, R. *et al.* (2014) 'Neural Oscillations and Synchronization Differentially Support Evidence Accumulation in Perceptual and Value-Based Decision Making', *Neuron*. Cell Press, 82(3), pp. 709-720. doi: 10.1016/j.neuron.2014.03.014.
- Raposo, D. *et al.* (2012) 'Multisensory decision-making in rats and humans', *Journal of Neuroscience*. Society for Neuroscience, 32(11), pp. 3726-3735. doi: 10.1523/JNEUROSCI.4998-11.2012.
- Ratcliff, R. (1978) 'A theory of memory retrieval', *Psychological Review*, 85(2), pp. 59-108. doi: 10.1037/0033-295X.85.2.59.
- Ratcliff, R. *et al.* (2016) 'Diffusion Decision Model: Current Issues and History', *Trends in Cognitive Sciences*. Elsevier Current Trends, 20(4), pp. 260-281. doi: 10.1016/j.tics.2016.01.007.
- Ratcliff, R., Cherian, A. and Segraves, M. (2003) 'A comparison of Macaque behavior and superior colliculus neuronal activity to predictions from models of two-choice decisions', *Journal of Neurophysiology*. Glimcher and Sparks, 90(3), pp. 1392-1407. doi: 10.1152/jn.01049.2002.
- Ratcliff, R. and Childers, R. (2015) 'Individual differences and fitting methods for the two-choice diffusion model of decision making', *Decision*. American Psychological Association Inc., 2(4), pp. 237-279. doi: 10.1037/dec0000030.
- Ratcliff, R. and Frank, M. J. (2012) 'Reinforcement-based decision making in corticostriatal circuits: Mutual constraints by neurocomputational and diffusion models', *Neural Computation*, 24(5), pp. 1186-1229. doi: 10.1162/NECO_a_00270.
- Ratcliff, R. and McKoon, G. (2008) 'The diffusion decision model: Theory and data for two-choice decision tasks', *Neural Computation*. MIT Press 238 Main St., Suite 500, Cambridge, MA 02142-1046 USA journals-info@mit.edu, 20(4), pp. 873-922. doi: 10.1162/neco.2008.12-06-420.
- Ratcliff, R., Philiastides, M. G. and Sajda, P. (2009) 'Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG',

Proceedings of the National Academy of Sciences. National Academy of Sciences, 106(16), pp. 6539-6544. doi: 10.1073/pnas.0812589106.

Ratcliff, R. and Smith, P. L. (2004) 'A Comparison of Sequential Sampling Models for Two-Choice Reaction Time', *Psychological Review*, 111(2), pp. 333-367. doi: 10.1037/0033-295X.111.2.333.

Ratcliff, R. and Tuerlinckx, F. (2002) 'Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability', *Psychonomic Bulletin and Review*. Springer-Verlag, 9(3), pp. 438-481. doi: 10.3758/BF03196302.

Regenbogen, C. *et al.* (2016) 'Bayesian-based integration of multisensory naturalistic perithreshold stimuli', *Neuropsychologia*. Elsevier Ltd, 88, pp. 123-130. doi: 10.1016/j.neuropsychologia.2015.12.017.

Reinagel, P. (2013) 'Speed and Accuracy of Visual Motion Discrimination by Rats', *PLoS ONE*. Edited by E. Arabzadeh. Public Library of Science, 8(6), p. e68505. doi: 10.1371/journal.pone.0068505.

Rohe, T., Ehrlis, A. C. and Noppeney, U. (2019) 'The neural dynamics of hierarchical Bayesian causal inference in multisensory perception', *Nature Communications*. Nature Publishing Group, 10(1), p. 1907. doi: 10.1038/s41467-019-09664-2.

Rohe, T. and Noppeney, U. (2015) 'Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception', *PLOS Biology*. Edited by C. Kayser. Public Library of Science, 13(2), p. e1002073. doi: 10.1371/journal.pbio.1002073.

Rohrbaugh, J. W., Donchin, E. and Eriksen, C. W. (1974) 'Decision making and the P300 component of the cortical evoked response', *Perception and Psychophysics*, 15(2), pp. 368-374. doi: 10.3758/BF03213960.

Romei, V. *et al.* (2009) 'Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds.', *Current biology: CB*, 19(21), pp. 1799-805. doi: 10.1016/j.cub.2009.09.027.

Rorie, A. E. and Newsome, W. T. (2005) 'A general mechanism for decision-making in the human brain?', *Trends in Cognitive Sciences*. Elsevier Current Trends, pp. 41-43. doi: 10.1016/j.tics.2004.12.007.

Rousselet, G. A., Foxe, J. J. and Bolam, J. P. (2016) 'A few simple steps to improve the description of group results in neuroscience', *European Journal of Neuroscience*. Wiley, 44(9), pp. 2647-2651. doi: 10.1111/ejn.13400.

Rousselet, G. A., Pernet, C. R. and Wilcox, R. R. (2017) 'Beyond differences in means: robust graphical methods to compare two groups in neuroscience', *European Journal of Neuroscience*, 46(2), pp. 1738-1748. doi: 10.1111/ejn.13610.

RStudio Team (2016) 'RStudio: integrated development for R', *RStudio, Inc., Boston, MA*,. Boston, MA, MA: RStudio, Inc.

Sajda, P., Philiastides, M. G. M. G. and Parra, L. C. (2009) 'Single-Trial Analysis of Neuroimaging Data: Inferring Neural Networks Underlying Perceptual Decision-Making in the Human Brain', *IEEE Reviews in Biomedical Engineering*, 2, pp. 97-109. doi: 10.1109/RBME.2009.2034535.

Schroeder, C. E. and Foxe, J. (2005) 'Multisensory contributions to low-level, "unisensory" processing', *Current Opinion in Neurobiology*. Elsevier Current Trends, 15(4), pp. 454-458. doi: 10.1016/j.conb.2005.06.008.

Shadlen, M. N. and Newsome, W. T. (1996) 'Motion perception: seeing and deciding.', *Proceedings of the National Academy of Sciences*, 93(2), pp. 628-633. doi: 10.1073/pnas.93.2.628.

Sheppard, J. P., Raposo, D. and Churchland, A. K. (2013) 'Dynamic weighting of multisensory stimuli shapes decision-making in rats and humans', *Journal of Vision*. The Association for Research in Vision and Ophthalmology, 13(6), pp. 4-4. doi: 10.1167/13.6.4.

Shi, Z. and Müller, H. J. (2013) 'Multisensory perception and action: development, decision-making, and neural mechanisms.', *Frontiers in*

integrative neuroscience, 7, p. 81. doi: 10.3389/fnint.2013.00081.

Siegel, M., Engel, A. K. and Donner, T. H. (2011) 'Cortical Network Dynamics of Perceptual Decision-Making in the Human Brain', *Frontiers in Human Neuroscience*. Frontiers Media S. A., 5(FEBRUARY), pp. 1-12. doi: 10.3389/fnhum.2011.00021.

Smith, P. L. and Little, D. R. (2018) 'Small is beautiful: In defense of the small-N design', *Psychonomic bulletin & review*. Springer, 25(6), pp. 2083-2101. doi: 10.3758/s13423-018-1451-8.

Spiegelhalter, D. J. *et al.* (2002) 'Bayesian measures of model complexity and fit', *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 64(4), pp. 583-616. doi: 10.1111/1467-9868.00353.

Stekelenburg, J. J. and Vroomen, J. (2007) 'Neural correlates of multisensory integration of ecologically valid audiovisual events', *Journal of Cognitive Neuroscience*. MIT Press 238 Main St., Suite 500, Cambridge, MA 02142-1046USA journals-info@mit.edu, 19(12), pp. 1964-1973. doi: 10.1162/jocn.2007.19.12.1964.

Stekelenburg, J. J. and Vroomen, J. (2012) 'Electrophysiological correlates of predictive coding of auditory location in the perception of natural audiovisual events.', *Frontiers in integrative neuroscience*, 6, p. 26. doi: 10.3389/fnint.2012.00026.

Tagliabue, C. F. *et al.* (2019) 'The EEG signature of sensory evidence accumulation during decision formation closely tracks subjective perceptual experience', *Scientific Reports*. Nature Publishing Group, 9(1), p. 4949. doi: 10.1038/s41598-019-41024-4.

Talsma, D. and Woldorff, M. G. (2005) 'Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity', *Journal of Cognitive Neuroscience*. MIT Press 238 Main St., Suite 500, Cambridge, MA 02142-1046 USA journals-info@mit.edu, 17(7), pp. 1098-1114. doi: 10.1162/0898929054475172.

The MathWorks, I. (2015a) 'MATLAB 2015a'. Natick, Massachusetts, United States: The MathWorks, Inc.

The MathWorks, I. (2015b) 'MATLAB and Statistics Toolbox Release 2015b'. Natick, Massachusetts, United States.

Tosoni, A. *et al.* (2008) 'Sensory-motor mechanisms in human parietal cortex underlie arbitrary visual decisions', *Nature neuroscience*. Nature Publishing Group, 11(12), pp. 1446-1453.

Troje, N. F. and Bühlhoff, H. H. (1996) 'Face recognition under varying poses: The role of texture and shape', *Vision Research*. Pergamon, 36(12), pp. 1761-1771. doi: 10.1016/0042-6989(95)00230-8.

Turner, B. M. *et al.* (2017) 'The dynamics of multimodal integration: The averaging diffusion model', *Psychonomic Bulletin and Review*. Springer New York LLC, 24(6), pp. 1819-1843. doi: 10.3758/s13423-017-1255-2.

Turner, B. M., Van Maanen, L. and Forstmann, B. U. (2015) 'Informing cognitive abstractions through neuroimaging: The neural drift diffusion model', *Psychological Review*. American Psychological Association Inc., 122(2), pp. 312-336. doi: 10.1037/a0038894.

Verdonck, S., Loossens, T. and Philiastides, M. G. (2020) 'The Leaky Integrating Threshold and Its Impact on Evidence Accumulation Models of Choice Response Time (RT)', *Psychological Review*. American Psychological Association Inc. doi: 10.1037/rev0000258.

Wabersich, D. and Vandekerckhove, J. (2014) 'Extending JAGS: A tutorial on adding custom distributions to JAGS (with a diffusion model example)', *Behavior Research Methods*. Springer, 46(1), pp. 15-28. doi: 10.3758/s13428-013-0369-3.

Wang, X. J. (2002) 'Probabilistic decision making by slow reverberation in cortical circuits', *Neuron*. Cell Press, 36(5), pp. 955-968. doi: 10.1016/S0896-6273(02)01092-9.

Wang, Y. *et al.* (2008) 'Visuo-auditory interactions in the primary visual cortex of the behaving monkey: Electrophysiological evidence', *BMC Neuroscience*. BioMed Central, 9(1), p. 79. doi: 10.1186/1471-2202-9-79.

Welford, W. T., Brebner, J. M. T. and Kirby, N. (1980) *Reaction times*. Stanford University.

Werner, S. and Noppeney, U. (2010a) 'Distinct Functional Contributions of Primary Sensory and Association Areas to Audiovisual Integration in Object Categorization', *Journal of Neuroscience*. Society for Neuroscience, 30(7), pp. 2662-75. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20164350> (Accessed: 24 May 2019).

Werner, S. and Noppeney, U. (2010b) 'Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization', *Cerebral Cortex*. Narnia, 20(8), pp. 1829-1842. doi: 10.1093/cercor/bhp248.

Wiecki, T. V., Sofer, I. and Frank, M. J. (2013) 'HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python', *Frontiers in Neuroinformatics*. Frontiers, 7, p. 14. doi: 10.3389/fninf.2013.00014.

Wu, Z. *et al.* (2020) 'Context-Dependent Decision Making in a Premotor Circuit', *Neuron*. Cell Press, 106(2), pp. 316-328.e6. doi: 10.1016/j.neuron.2020.01.034.

Wyart, V. *et al.* (2012) 'Rhythmic Fluctuations in Evidence Accumulation during Decision Making in the Human Brain', *Neuron*. Cell Press, 76(4), pp. 847-858. doi: 10.1016/j.neuron.2012.09.015.